THE UNIVERSITY OF CHICAGO

SEARCH FOR $ZW/ZZ \to \ell^+\ell^-$ + JETS PRODUCTION IN $p\overline{p}$ COLLISIONS AT CDF

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF PHYSICS

BY
WESLEY KETCHUM

CHICAGO, ILLINOIS
DECEMBER 2012

ABSTRACT

The Standard Model of particle physics describes weak interactions mediated by massive gauge bosons that interact with each other in well-defined ways. Observations of the production and decay of WW, WZ, and ZZ boson pairs are an opportunity to check that these self-interactions agree with the Standard Model predictions. Furthermore, final states that include quarks are very similar to the most prominent final state of Higgs bosons produced in association with a W or Z boson. Diboson production where WW is a significant component has been observed at the Tevatron collider in semi-hadronic decay modes. We present a search for ZW and ZZ production in a final state containing two charged leptons and two jets using 8.9 fb^{-1} of data recorded with the CDF detector at the Tevatron. We select events by identifying those that contain two charged leptons, two hadronic jets, and low transverse missing energy ($\not\!\!E_T$). We increase our acceptance by using a wide suite of high-p_T lepton triggers and by relaxing many lepton identification requirements. We develop a new method for calculating corrections to jet energies based on whether the originating parton was a quark or gluon to improve the agreement between data and the Monte Carlo simulations used to model our diboson signal and dominant backgrounds. We also make use of neural-network-based discriminants that are trained to pick out jets originating from b quarks and light-flavor quarks, thereby increasing our sensitivity to $Z \to b\bar{b}$ and $W/Z \to q\bar{q}'$ decays, respectively. The number of signal events is extracted through a simultaneous fit to the dijet mass spectrum in three channels: a heavy-flavor tagged channel, a light-flavor tagged channel, and an untagged channel. We measure $\sigma_{ZW/ZZ} = 2.5^{+2.0}_{-1.0}$ pb, which is consistent with the SM cross section of 5.1 pb. We establish an upper limit on the cross section of $\sigma_{ZW/ZZ} < 6.1$ pb at 95% CL.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

There are many people I would like to thank, but first and foremost I would like to thank my advisor, Young-Kee Kim. Young-Kee, despite the impossibility of tracking you down sometimes, I always felt like you were there and helping me. I am greatly indebted to you for too many things: for your insight and help in focusing my work; for the times you have pushed me to finish rather than let me circle endlessly around a problem; for the trust you placed in me to mentor students; for the confidence you placed in me as I applied for jobs; and, for the way you invited Elizabeth and I into your home and life. You and Sid are models for the type of person I hope to be: kind, energetic, and endlessly curious. While it seems crazy to say, I will genuinely miss our Sunday morning meetings. I look forward to remain working at Fermilab for many reasons, but a big one is that I will still be around to see you. Who knows—perhaps I'll see you even more often!

I owe another gigantic portion of gratitude to my other advisor, Vadim Rusu. Vadim, it's hard to express how much help you've given me over these past five years. You pushed me to think bigger and do more than I could have ever done on my own. Your honesty helped me be critical of my own work, but when I was at my least confident, you always seemed to be there to back me up. You were a spectacular teacher and mentor, and I can't be more thrilled that we'll likely see even more of each other.

There were many other people I had the pleasure to work closely with, especially Zaid, Stephen, Marco, Sasha, John, and Peter. You all contributed integral parts to my own work, and provided a community of fellow researchers that made it fun. I also thank the other REU and undergraduate students I've worked with over the years (Jake, Michael, Lauren, James, Rachel, David, and Sid...though I'm sure I'm forgetting some): I learned

far more than you possibly could have learned from me.

There were also many people at CDF, particularly the Higgs group conveners (Ben, Eric, Craig, and Homer) that have helped me to make my work even better. One of the few downsides to spending a lot of time at the university and in the city was that I didn't get to spend more time getting to know many at CDF better, but I am happy that I may yet get that chance.

I very much appreciate all the people at the university who made the environment there what it was, particularly to the people in the CDF group (Henry, Mel, Carla, Florencia, Kohei, Wojtek, Jahred, Erik, Shawn, Scott, Dan, Martina, Hyunsu, and Jian). And I especially thank all of my classmates, who I could not only turn toward for help when I needed it most, but who I hope will be lifelong friends.

Finally, I want to thank my family, who have always looked out for me and cheered me on, and feigned interest in the details of my work. Most importantly, I want to thank my wife and best friend, Elizabeth. You have always been there to help and keep me calm, to sympathize and empathize, and to share in the happy and sad moments. If I have grown at all, it is because of you.

CHAPTER 1

INTRODUCTION

1.1 Fundamental Elements of Matter

The Standard Model (SM) of particle physics describes the interactions of the fundamental particles that make up the known matter in our universe. The basic constituents of the theory include twelve fermions and their twelve antiparticle counterparts. The fermions are split evenly between quarks—those that have "color" and so interact with the strong force—and the leptons, which do not. They are also split into three "generations": the first generation contains the up and down quarks, along with the electron and its associated neutrino; the heavier second generation contains the strange and charm quarks and the muon, along with its neutrino; lastly, the even more massive third generation contains the tau lepton (and tau neutrino) with the bottom and top quark—the latter is the most massive elementary particle known. The matter most familiar in everyday life is made up of particles in the first generation; up and down quarks combine to form the protons and neutrons of atomic nuclei, and those nuclei combine with electrons to form atoms.

The Standard Model describes interactions of these fundamental particles under three forces. The most familiar is the electromagnetic (EM) force, mediated by the photon which couples with charged particles and their antiparticles. A formal description of the electromagnetic force in the language of quantum field theory is given by quantum electrodynamics (QED); the EM force is described as a gauge theory with the symmetry group U(1). QED successfully describes a large number of phenomona related to electromagnetism.

The strong nuclear force is described in the Standard Model through the theory of

Figure 1.1: A summary of the particles included in the Standard Model of particle physics, and where and when they were discovered. The figure includes the quarks (red), charged leptons (yellow), neutrinos (green), and force mediators (blue/purple). Not shown is the Higgs boson, recently discovered at CERN. Figure adapted from [1].

quantum chromodynamics (QCD). Quarks carry a “color charge” (typically given values of red, green, or blue) analogous to the electric charge in electromagnetism. They interact with gluons, the mediators of the strong force. The strength of the strong force, unlike the electromagnetic force, does not increase in interactions at higher energies and shorter distances. Instead, the QCD exhibits asymptotic freedom. The strong force behaves similar to a spring or elastic band with a coupling that increases at longer distances. This behavior is responsible for holding nucleons together, but also prevents free quarks from ordinarily appearing in nature; rather, quarks typically appear in groups that together form a color neutral state. Particles made up of quarks are called hadrons—those made up of one quark and one antiquark (*e.g.*, the pions and kaons) are called mesons, while those made up of three quarks or three antiquarks (like the proton, antiproton, and neutron) are called baryons.

Finally, the Standard Model includes interactions that describe the weak nuclear force, responsible for nuclear β decays and other phenomena. The two quarks and two leptons in each generation make up a doublet of in SU(2). Weak interactions between the particles in these doublets ares mediated by massive $W^{\pm}$ and Z bosons. The Standard Model is able to incorporate many of the more shocking properties of the weak force, like that it does not respect parity (P) or charge-parity (CP) symmetries. The weak force and how it is combined with the electromagnetic force to form a unified “electroweak” theory is described in greater detail in the following sections.

1.1.1 The Weak Interaction in Detail

The study of the weak interaction has its beginning in the observations of radioactive decay near the end of the 19$^{\text{th}}$ century. Work by Ernest Rutherford and Henri Becquerel established that, among the types of radioactive decays discovered, there was the emis-

sion of a particle that had the same charge to mass ratio as the electron. The elements of this β-radiation were thus identified as being electrons themselves. Further study of the electrons from β decays showed that their momentum distribution was continuous rather than discrete. This led Wolfgang Pauli to postulate that an additional particle was produced in β decays: the neutrino. Meanwhile, work showed that atomic nuclei undergoing β decays increased their atomic number by +1.

The final piece necessary to develop a complete model for nuclear β decay similar to how we know it today was the discovery of the neutron by J. J. Chadwick in 1932. With the interacting particles identified, Enrico Fermi developed a theory describing β-radiation as the decay of the neutron: $n \to p^+ e^- \bar{\nu}_e$. Fermi's interaction, shown in Fig. 1.2, described the destruction of a neutron and creation of an electron, proton, and antineutrino at the same point in space-time. While Fermi's original interaction conserved the angular momentum of the nucleus and parity, changes were made to accommodate experimental results that showed these quantities were not conserved in β decays. Muon, charged pion, and eventually strange hadron decays were all identified as following similar behavior, and so these were all successfully incorporated into Fermi's four-point weak interaction theory.

However, Fermi's theory predicted unphysical cross sections at higher energies, as the cross sections increased quickly with the center-of-mass energy. These cross sections could be held under control if the weak interaction was mediated by a vector boson (the $W^{\pm}$) that carried electric charge. However, this boson would have to be massive in order to explain the short range of the weak force as compared to the electromagnetic force, and a theory with a massive boson mediator is not locally gauge invariant. This problem can be solved via the Higgs mechanism [2, 3, 4], described in Sec. 1.1.2, which allows for the generation of massive W bosons in a locally gauge invariant theory with a spontaneously broken symmetry. A more complete theory that combined the electromagnetic and weak

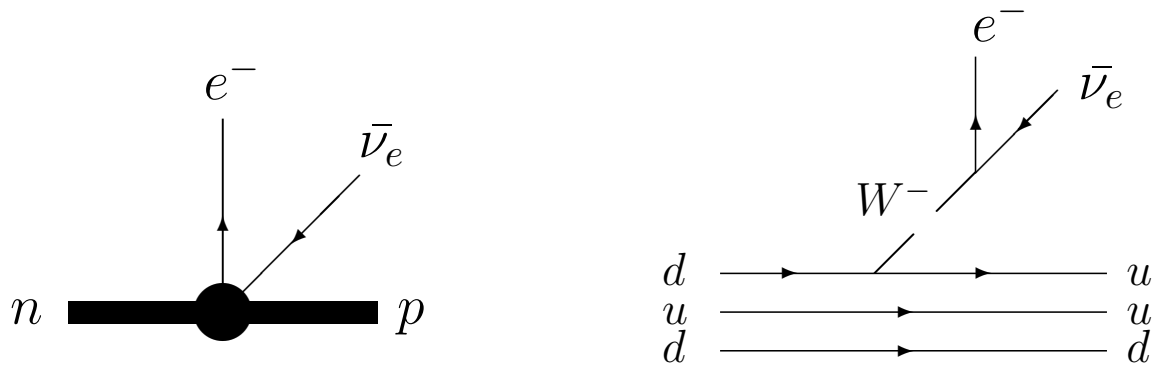

Figure 1.2: Feynman diagrams for neutron decays describing the weak interaction in Fermi's four-point interaction theory (left) and the electroweak theory where the force is mediated by the massive *W* boson (right). We show the quarks that make up the neutron and proton, as it is the quarks that undergo the interaction.

interactions using the existence of a universal Higgs field was developed by Glashow, Weinberg, and Salam [5, 6, 7].

In this theory, we end up with massive W^+ and W^- bosons, which mediate the weak interactions described by Fermi's theory. A diagramatic description of nuclear β decay in both Fermi's original four-point interaction and the interaction with a massive *W* is shown in Fig. 1.2. In addition to a massive *W* boson, the GSW theory predicts the existence of the massive, neutral *Z* boson, which mediates neutral-weak-current interactions. Observations of such interactions with neutrinos in the 1970s provided early indirect evidence for the *Z* boson [8], and both the *W* and *Z* bosons were discovered directly in the early 1980s by the UA-1 [9, 10] and UA-2 [11, 12] collaborations at the Sp$\bar{\text{p}}$S collider at CERN.

1.1.2 Electroweak Symmetry Breaking

As mentioned in Sec. 1.1.1, it is possible to combine the weak and electromagnetic forces in a single theory based on the group SU(2) $\otimes$ U(1). We start with three gauge bosons

associated with the $SU(2)$ symmetry, A^a ($a = 1, 2, 3$), and one with the U(1) symmetry, B. The free Lagrangian for these bosons is

$$\mathcal{L} = -\frac{1}{4}A^{a,\mu\nu}A^a_{\mu\nu} - \frac{1}{4}B^{\mu\nu}B_{\mu\nu} - \frac{1}{2}m^2_{A^a}A^{i,\mu}A^i_\mu - \frac{1}{2}m^2_B B^\mu B_\mu \ , \tag{1.1}$$

where a sum over a is implied, and where $B^{\mu\nu} \equiv \partial^\mu B^\nu - \partial^\nu B^\mu$ with a similar definition for $A^{a,\mu\nu}$. However, the latter two mass terms in Eq. 1.1 are not locally gauge invariant, and so in order to preserve gauge invariance in this SU(2) $\otimes$ U(1) theory, we must take the A^a and B bosons to be massless.

In order to generate a mass for these bosons, we introduce a scalar doublet ϕ in SU(2),

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}$$

that has a "hypercharge" of $+\frac{1}{2}$ under the U(1) symmetry. Additionally, we introduce a potential to the Lagrangian of the form

$$V(\phi) = -\mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2 \ ,$$

where μ and λ are constants. The Lagrangian describing our bosons and interactions with this new scalar field becomes

$$\begin{aligned}\mathcal{L} &= -\frac{1}{4}A^{a,\mu\nu}A^a_{\mu\nu} - \frac{1}{4}B^{\mu\nu}B_{\mu\nu} + D^\mu\phi^\dagger D_\mu\phi - V(\phi) \\ &= -\frac{1}{4}A^{a,\mu\nu}A^a_{\mu\nu} - \frac{1}{4}B^{\mu\nu}B_{\mu\nu} + D^\mu\phi^\dagger D_\mu\phi + \mu^2\phi^\dagger\phi - \lambda(\phi^\dagger\phi)^2 \ .\end{aligned}$$

The covariant derivative is

$$D_\mu\phi = (\partial_\mu - i\frac{1}{2}A^a_\mu \cdot \sigma^a - 1\frac{g'}{2}B_\mu)\phi \ ,$$

where g and g' represent field strengths of the A^a and B fields, and σ^a is the a^{th} Pauli matrix.

The structure of the potential $V(\phi)$ is symmetric under SU(2) $\otimes$ U(1) gauge transformations of ϕ, and has a minimum away from $|\phi| = 0$ at

$$|\phi_{\text{min}}| = \sqrt{\frac{\mu^2}{2\lambda}} \equiv \frac{v}{\sqrt{2}} \, .$$

The Lagrangian contains only terms that are gauge-invariant, but the potential has a mechanism by which ϕ will move towards a non-zero value. Though it may reach a minimum by moving in any direction, we break the SU(2) symmetry by choosing a particular direction in SU(2) for the location of that minimum. We say the field acquires a vacuum expectation value of

$$\langle \phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \, .$$

Evaluating the covariant derivative term in the Lagrangian at this expectation value of ϕ, we get new terms in the Lagrangian that are quadratic in the boson fields:

$$\Delta\mathcal{L} = \frac{1}{2}\frac{v^2}{4}\left[g^2(A^{1,\mu}A^1_\mu) + g^2(A^{2,\mu}A^2_\mu) + (gA^{3,\mu} - g'B^\mu)(gA^3_\mu - g'B_\mu)\right] \, . \tag{1.2}$$

We then make the following rearrangement in terms of new fields, W^+_μ, W^-_μ, Z^0_μ, and A_μ:

$$A^1_\mu = \frac{1}{\sqrt{2}}(W^+_\mu + iW^-_\mu) \, ;$$

$$A^2_\mu = \frac{1}{\sqrt{2}}(W^+_\mu - iW^-_\mu) \, ;$$

$$A^3_\mu = \frac{1}{\sqrt{g^2 + g'^2}}(gZ^0_\mu + g'A_\mu) \, ;$$

$$B_\mu = \frac{1}{\sqrt{g^2 + g'^2}}(-g' Z^0_\mu + g A_\mu) \ .$$

Substituting these into Eq. 1.2, we find

$$\Delta\mathcal{L} = \frac{1}{2}\frac{v^2}{4}\left[g^2(W^{+,\mu}W^+_\mu) + g^2(W^{-,\mu}W^-_\mu) + (g^2 + g'^2)Z^{0,\mu}Z^0_\mu\right] \ ,$$

which are mass terms for the two W bosons, with $m_W = \left(\frac{v}{2}\right) g$, and the Z boson, with mass $m_Z = \left(\frac{v}{2}\right)\sqrt{g^2 + g'^2}$. The A_μ field, describing the photon in the Standard Model, remains massless. One may interpret the two neutral bosons in the original theory as having mixed to form the massive Z boson and the massless γ with a mixing matrix,

$$\begin{pmatrix} A^3_\mu \\ B_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} Z^0_\mu \\ A_\mu \end{pmatrix}$$

where the mixing angle θ_W follows

$$\sin^2\theta_W = \frac{g'^2}{\sqrt{g^2 + g'^2}} \ .$$

This mixing angle may be treated as a fundamental input into the electroweak theory, as it describes the relative interaction strengths of the electromagnetic and weak forces. It may also be used to describe the relationship between the masses of the W and Z bosons: $m_W = m_Z \cos\theta_W$. The electric charge (Q) is a mixture of the weak isospin (T_3) and weak hypercharge (Y):

$$Q = T_3 + \frac{Y}{2}$$

While discovery of the W and Z bosons confirmed many aspects of this theory, direct experimental evidence of the Higgs boson alluded physicists until very recently. This is in part due to the fact that the mass of the Higgs boson, given by $m_H = v\sqrt{2\lambda} = \mu$,

is not predicted by electroweak theory and instead must be measured experimentally. Indirect evidence from combining other measurements of fundamental parameters in the Standard Model suggested that the Higgs boson mass was in the 100 GeV/c^2 range, with $m_H = 96^{+31}_{-24}$ GeV/c^2 [13]. Direct searches at LEP, the e^+e^- collider at CERN that ran from 1989 until 2000, excluded a Higgs boson with mass < 114.4 GeV/c^2 [14], suggesting there may be tensions with the SM. However, in 2012, the ATLAS and CMS experiments located at CERN's Large Hadron Collider (LHC) reported observations of a new Higgs-like particle with $m_H \sim 125$ GeV/c^2 [15, 16]. Additionally, the CDF and D0 experiments located at Fermilab's Tevatron found evidence consistent with the LHC discovery [17]. Further study is necessary to prove this particle is the SM Higgs boson, but early indications are consistent with it being so.

1.1.3 Decays of the W and Z bosons

The W and Z bosons interact with fermions in isospin doublets, as described in Sec. 1.1. The coupling of the W boson to left-handed quarks and leptons is similar (except for the flavor-changing properties introduced via quark-mixing, but this may be ignored when integrating over all quark states). Because quarks have color, however, there is an additional color factor that makes $W^{\pm} \to q\bar{q}'$ decays more likely. A W^+ boson may decay to one of 6 possible quark final states (there are no top quark final states because $m_t > m_W$):

$$\{u_r, \bar{d}_r\}, \{u_b, \bar{d}_b\}, \{u_g, \bar{d}_g\}, \{c_r, \bar{s}_r\}, \{c_b, \bar{s}_b\}, \{c_g, \bar{s}_g\}$$

and 3 possible lepton final states

$$\{e^+, \nu_e\}, \{\mu^+, \nu_\mu\}, \{\tau^+, \nu_\tau\}$$

with equal probability. Therefore, the branching ratio for each individual lepton decay is roughly $\mathcal{BR}(W^+ \to \ell^+\nu_\ell) \approx \frac{1}{9} \approx 11.1\%$, while the decay to any final state with quarks has $\mathcal{BR}(W^+ \to q\bar{q}') \approx \frac{6}{9} \approx 66.7\%$.

The couplings of the Z boson to fermions are more complicated due to the mixing of neutral boson states that creates it during spontaneous symmetry breaking. They are proportional to

$$T_3 - Q\sin^2\theta_W \,..$$

Unlike the W boson, the Z boson couples to both left- and right-handed fermions, but it does so differently due to parity violation in the weak force. Thus, to calculate the decay width of Z bosons to a fermion pair, we must add up both left and right-handed fermions interactions. The partial decay width for a Z boson to a charged lepton pair is

$$\Gamma_{Z\to\ell^+\ell^-} \propto \left[\left(-\frac{1}{2}\right) - (-1)\sin^2\theta_W\right]^2 + \left[(0) - (-1)\sin^2\theta_W\right]^2 \approx 0.13\ ,$$

while for the decay to an up- or down-type quarks pair, the decay width goes as

$$\Gamma_{Z\to q_{uc}\bar{q}_{uc}} \propto \left[\left(+\frac{1}{2}\right) - \left(+\frac{2}{3}\right)\sin^2\theta_W\right]^2 + \left[(0) - \left(+\frac{2}{3}\right)\sin^2\theta_W\right]^2 \approx 0.14\ ,$$

$$\Gamma_{Z\to q_{dsb}\bar{q}_{dsb}} \propto \left[\left(-\frac{1}{2}\right) - \left(-\frac{1}{3}\right)\sin^2\theta_W\right]^2 + \left[(0) - \left(-\frac{1}{3}\right)\sin^2\theta_W\right]^2 \approx 0.19\ ,$$

and for neutrinos, which in the Standard Model are only left-handed, the decay width will go as

$$\Gamma_{Z\to\nu_\ell\bar{\nu}_\ell} \propto \left[\left(+\frac{1}{2}\right) - (0)\sin^2\theta_W\right]^2 = 0.25\ ,$$

using $\sin^2\theta_W = 0.23$ [18]. Once again, we must consider the color possibilities for the quarks, as each type may be of three different colors, which leads to the following branch-

ing ratios

$$\mathcal{BR}(Z \to e^+e^-) = \mathcal{BR}(Z \to \mu^+\mu^-) = \mathcal{BR}(Z \to \tau^+\tau^-) = \frac{\Gamma_{Z\to\ell^+\ell^-}}{\Gamma_{Z,\text{Total}}} \sim \frac{0.13}{3.69} \approx 3.4\% \,,$$

$$\mathcal{BR}(Z \to q\bar{q}) = \frac{3 \times (2\Gamma_{Z\to q_{uc}\bar{q}_{uc}} + 3\Gamma_{Z\to q_{dsb}\bar{q}_{dsb}})}{\Gamma_{Z,\text{Total}}} \sim \frac{2.55}{3.69} \approx 69\% \,,$$

$$\mathcal{BR}(Z \to \nu\bar{\nu}(\text{all flavors})) = \frac{3 \times \Gamma_{Z\to\nu_\ell\bar{\nu}_\ell}}{\Gamma_{Z,\text{Total}}} \sim \frac{0.75}{3.69} \approx 20\% \,,$$

where $\Gamma_{Z,\text{Total}} \approx 3.69$ is the sum of all of the decay widths of the Z boson. The branching ratios for Z and W boson decays are shown in Fig. 1.3.

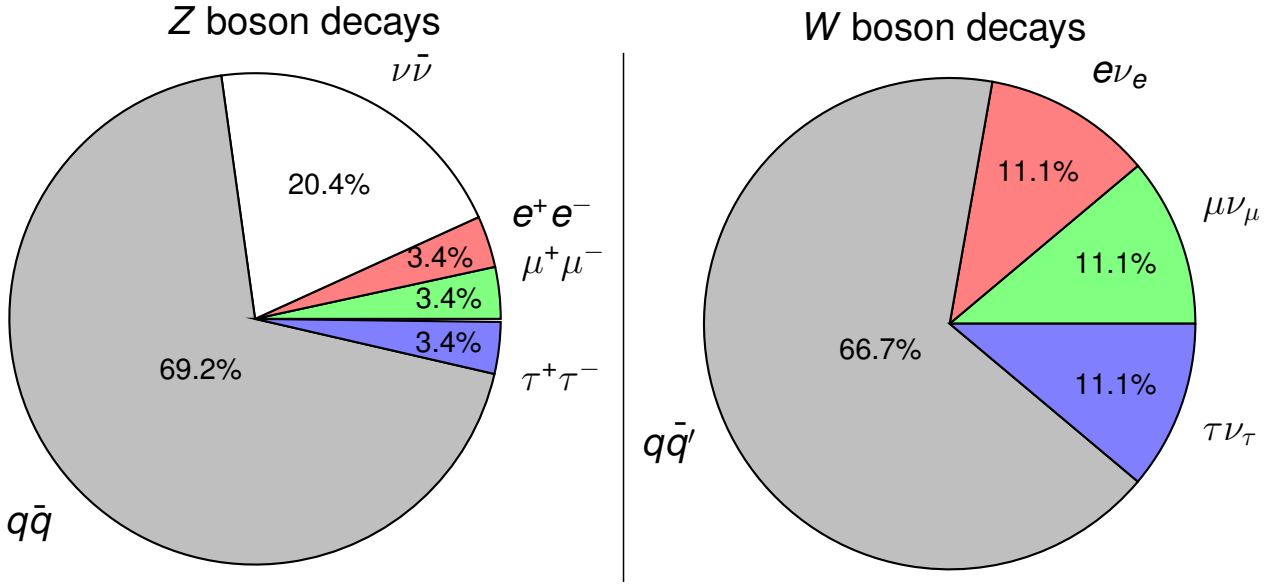

Figure 1.3: The branching ratios for Z boson decays (left) and W boson decays (right). The branching ratio for $Z \to b\bar{b}$ decays is $\approx \frac{3\times 0.19}{3.69} \approx 15\%$.

1.2 Diboson Production

There are a number of reasons why we are interested in studying the production of massive boson pairs. First, we may search for physics beyond the Standard Model in diboson

production. The Standard Model makes precise predictions to the self-couplings of the electroweak gauge bosons (the W and Z bosons and γ) [19]. Trilinear couplings with $WW\gamma$ and WWZ vertices are allowed in the SM at tree-level, while $ZZ\gamma$ and ZZZ couplings do not exist. New physics may change these couplings above some energy scale, either in modes allowed by the Standard Model or in those not present [20, 21].

At hadron colliders, the best limits on anomalous trilinear gauge couplings are typically achieved by looking at fully leptonic decay modes: $W^+W^- \to \ell^+\nu\ell^-\bar{\nu}$, $W^\pm Z \to \ell^\pm\nu\ell^+\ell^-$, and $ZZ \to \ell^+\ell^-\ell^+\ell^-$ (or $\ell^+\ell^-\nu\bar{\nu}$). These decay modes, while suffering from the low branching rations of leptonic W and Z decays, have few backgrounds due to their lack of jets in the final states. However, it is possible that new physics may be hiding in decays with hadronic final states [22, 23]. An excess of events observed by the CDF collaboration in a $W \to \ell\nu$ + jets final state [24] could be explained by such new physics, though this observation has not been corroborated by other experiments [25, 26, 27].

Second, searches for diboson production are strongly related to searches for the Higgs boson. For a Higgs boson at the masses indicated by the ATLAS [15] and CMS [16] experiments, $m_H \sim 125$ GeV/c^2, the most likely Higgs decay product is a $b\bar{b}$ pair, as shown in Fig. 1.4. However, $H \to b\bar{b}$ decays are practically impossible to observe on their own, as the backgrounds from b jet production via gluon splitting at hadron colliders are far too large, not to mention the added difficulty of picking out b jets from the even larger multijet production backgrounds. Among the best prospects for observing $H \to b\bar{b}$ decays comes when the Higgs boson is produced in association with a W or Z boson. If that W or Z boson decays leptonically, the signature from that decay may be used to pre-select events, and then a requirement of two b jets may be used to search for the Higgs boson. These "associated production" modes with $WH \to \ell\nu b\bar{b}$ and $ZH \to \ell^+\ell^-$(or $\nu\bar{\nu}$) $b\bar{b}$ have produced the best results on searches for $H \to b\bar{b}$ [17], and may be essential to show that the Higgs-like signature seen at the LHC experiments is indeed the Higgs boson.

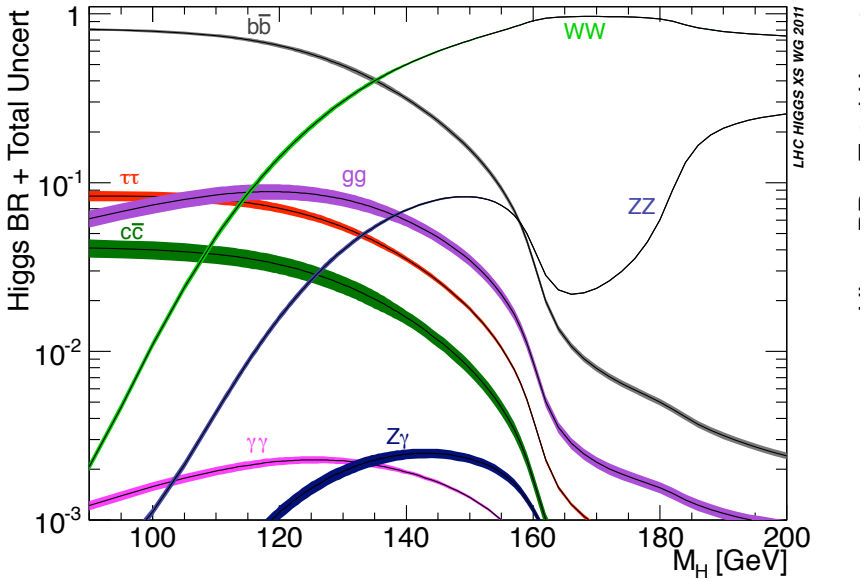

Figure 1.4: The branching ratios for various decays of the Higgs Boson as a function of m_H. The shaded regions indicate theoretical uncertainties. For $m_H \approx 125$ GeV/c^2, the most likely Higgs boson decay product is a $b\bar{b}$ pair. Associated Higgs production with a leptonic W or Z decay may allow for observations of $H \rightarrow b\bar{b}$ decays. Figure adapted from [28].

Electroweak diboson production is important to searches for the Higgs boson in associated production modes because semi-hadronic diboson decay channels, where one W or Z boson decays leptonically and the other decays hadronically, result in a similar final state: leptons plus jets from the decay of a massive boson. The similarities can be made even more germane when searching exclusively for $Z \to b\bar{b}$ decays in association with a leptonic W or Z boson decay. Here, the topology of the final states for diboson production and associated Higgs production differ primarily in the mass and spin of the Z and Higgs boson, but little else. Because production cross sections are larger for dibosons than the Higgs, searches for diboson production allow for the testing of new tools aimed at going after Higgs production. Sharing a final state means they also share many backgrounds, and so the detailed understanding of the W/Z+ jets and multijet production backgrounds from diboson production measurements are directly relevant to searches for $H \to b\bar{b}$ decays.

1.2.1 Existing Measurements

A recent review of diboson production measurements from the LEP, Tevatron, and LHC experiments may be found in [29]. Here, we will focus on observations of diboson production at hadron colliders in semi-hadronic decay modes, where one boson decays hadronically while the other decays leptonically. Production cross sections for heavy boson pairs at the Tevatron [19] and the LHC [30] are summarized in Tab. 1.1. These cross sections are calculated using the parton-level next-to-leading order program MCFM (described in [19]). Because $WW\gamma$ and WWZ couplings exist in the Standard Model, WW and WZ boson pairs may be produced through s-channel production as well as t-channel production (see Fig. 1.5). These additional s-channel processes, along with the lower W boson mass, give rise to the heirarchy of diboson cross sections. W^+W^- production is by far the

most prominent, with a cross section over twice that of the other processes at the Tevatron. ZZ production has a small cross section, as it may only be produced via t-channel exchange of a quark.

Diboson Cross Sections (pb)				
	Tevatron ($p\bar{p}$)	LHC (pp)		
	$\sqrt{s}$ = 1.96 TeV	$\sqrt{s}$ = 7 TeV	$\sqrt{s}$ = 8 TeV	$\sqrt{s}$ = 14 TeV
W^+W^-	11.7	47.0	57.3	124.3
W^+Z	3.7	11.9	14.5	31.5
W^-Z		6.7	8.4	20.3
ZZ	1.4	6.5	7.9	17.7

Table 1.1: A summary of the Standard Model production cross sections for heavy diboson production at the Tevatron [19] and at various center-of-mass energies at the LHC [30]. The W^+Z and W^-Z cross sections are combined for the Tevatron since they are equal to one another (due to the CP-symmetric initial $p\bar{p}$ state).

CDF made the first observation of diboson production in a semi-hadronic final state by searching for events with jets and a large imbalance in the transverse momentum [31]. The large momentum imbalance is caused by particles produced in $p\bar{p}$ collisions that escape detection: mainly neutrinos. Thus, the final state has a large acceptance to diboson decays as it includes contributions from both $W \to \ell\nu$ and $Z \to \nu\bar{\nu}$ decays. This increased acceptance increases sensitivity, but the analysis demands a good understanding of both W/Z+ jets backgrounds and the large number of events where mismeasured jet energies fake the neutrino signature. A combined measurement of WW, WZ, and ZZ production yielded $\sigma_{WW+WZ+ZZ} = 18.0 \pm 3.8$ pb, a significance in excess of 5σ and in agreement with the SM prediction. A later search [32] in the same final state included a channel designed to select events with $Z \to b\bar{b}$ decays, shifting the measurement to just the $WZ + ZZ$ production cross section. That search obtained a measurement of $\sigma_{WZ+ZZ} = 5.0^{+3.6}_{-3.0}$ pb, setting an upper limit of $\sigma_{WZ+ZZ} < 13$ pb at 95% C.L.

Measurements of final states requiring a $W \to \ell\nu$ decay by explicitly searching for a

charged lepton do not face as many backgrounds as the above analysis, but they lose contributions from $Z \to \nu\bar{\nu}$ decays, and so they are only sensitive to a combination of *WW* and *WZ* production. Both CDF [33] and D0 [34] have conclusively observed and measured the *WW* + *WZ* cross section in this final state, using methods ranging from a fit to the dijet mass distribution to more sophisticated multi-variate analysis techniques. Both analyses used an integrated luminosity of $\int \mathcal{L}$ = 4.3 fb^{-1}: CDF found σ_{WW+WZ} = 18.1 $\pm$ 4.1 pb, while D0 measured $\sigma_{WW+WZ} = 19.6^{+3.2}_{-3.0}$ pb. In addition to the combined measurement, D0 also used *b*-jet tagging to try to isolate the *WW* and *WZ* samples from one another, finding $\sigma_{WW} = 15.9^{+3.7}_{-3.2}$ pb and $\sigma_{WZ} = 3.3^{4.1}_{-3.3}$ pb. The measurements with *WW* contributions in this final state are all higher, but still consistent with, the SM prediction.

In addition to these dedicated diboson production measurements, both CDF and D0 have produced evidence of *WZ* and *ZZ* production using analysis techniques specifically developed for Higgs searches in the *WH* and *ZH* associated production modes. CDF, combining channels utilizing a variety of the leptonic decays of the *W* and *Z* bosons in concert with a $Z \to b\bar{b}$ decay, found $\sigma_{WZ+ZZ} = 4.1 \pm 1.3$ pb, corresponding to a significance of $\sim 3.2\sigma$ [35]. D0, combining similar channels, found σ_{WZ+ZZ} = 3.3 $\pm$ 1.4 pb, with a statistical significance of $\sim 2.5\sigma$ [36]. Taken together, the combined Tevatron measurement for WZ/ZZ production in leptons + *b* jets final states is σ_{WZ+ZZ} = 3.9 $\pm$ 0.9 pb [17].

1.3 Search for ZW/ZZ Production at CDF

This thesis is a search for ZW/ZZ production in a final state with two charged leptons and at least two jets: $ZW/ZZ \to \ell^+\ell^-$ + $q\bar{q}'$. Feynman diagrams describing the tree-level production and decay modes are shown in Fig. 1.5. This channel has a number of benefits compared to other semi-hadronic diboson production searches. First, high-

momentum electrons/positions and muons/antimuons generally leave unique signatures in the detector and are among the easiest particles to distinguish from hadronic jets. At a proton collider, where the total $p\overline{p}$ (or pp) interaction cross section is dominated by the production of hadronic jets, using a final state with charged leptons reduces these multijet production backgrounds significantly.

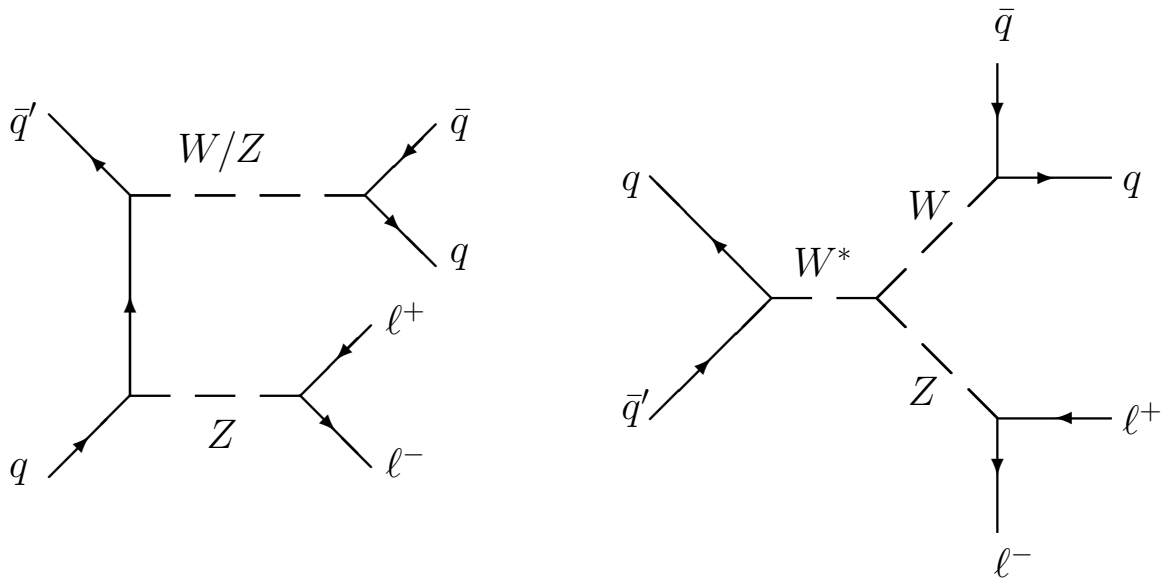

Figure 1.5: The leading-order Feynman diagrams for *ZW* and *ZZ* production at the Tevatron. The *t*-channel production (left) can produce both *ZW* and *ZZ* diboson final states, while *ZW* may also be produced via the *s*-channel production of an off-shell *W* boson (right).

In addition to the clean signature of electrons and muons, we may additionally reduce possible backgrounds by utilizing the fact that the reconstructed dilepton invariant mass should produce a resonance at the *Z* boson mass. By requiring the dilepton invariant mass to be in a close window around m_Z, we almost entirely reduce our backgrounds to processes with $Z \to \ell^+\ell^-$ decays themselves. Also, each product in our final state should be reconstructed in the CDF detector. Thus, we can place cuts to remove both physical processes that produce neutrinos (like $t\bar{t}$ production), as well as events where

some objects are highly mismeasured, which could lead to errors in event reconstruction.

However, this final state does have a significant drawback. As shown in Sec. 1.1.3, the branching ratio of $Z \to e^+e^-$ and $\mu^+\mu^-$ decays is very low. Only 4.5% of WZ and 9.4% of ZZ decays end in our search channel, reducing the combined cross section from ~ 5.1 pb to an effective cross section of $\sigma_{WZ+ZZ} \times \mathcal{BR}(ZW/ZZ \to \ell^+\ell^- q\bar{q}') \sim 0.30$ pb.

We perform our search for ZW/ZZ production using the full set of data produced by the Tevatron, corresponding to an integrated luminosity of 8.9 fb^{-1}. We attempt to extract our diboson signal from the backgrounds in our selection in a relatively simple way, by performing a fit to the dijet invariant mass (m_{jj}) spectrum. In order to increase our sensitivity to our signal, we separate events that pass our final selection into three channels: a heavy-flavor tagged channel, richer in jets originating from b quarks; a light-flavor tagged channel, richer in jets originating from light-flavor quarks (as opposed to gluons); and, a third "untagged" channel containing the remaining events. These two jet taggers, described in Chaps. 6 and 7, are both based on artificial neural-network discriminants that have been specifically designed with performing diboson searches in mind, the latter quark/gluon discriminant is unique to this analysis. We perform the fit to m_{jj} simultaneously across all channels.

We begin this thesis by describing the Tevatron accelerator and CDF detector in Chap. 2. How we use signals from the detector to reconstruct fundamental objects—like electrons, muons, and jets—is described in Chap. 3. We summarize details of the event selection and modeling of our signal and backgrounds on Chap. 4 before returning to a detailed description of the calibrations and corrections applied to our jet energies in Chap. 5. That latter chapter includes a description of new corrections applied to simulated jets based on their originating parton. Chaps. 6 and 7 describe the neural-network based b-jet and quark/gluon discriminants that we use to improve our sensitivity to $Z \to b\bar{b}$ and $W/Z \to q\bar{q}'$ decays. Finally, in Chap. 8, we put everything together to perform our fit to

the m_{jj} spectrum and describe the systematic uncertainties on the signal extraction.

CHAPTER 2

ACCELERATOR AND DETECTOR

2.1 The Tevatron Accelerator Chain

The Tevatron, located at the Fermi National Accelerator Laboratory (FNAL, or Fermilab), accelerated protons and antiprotons for use in high energy particle physics experiments from 1983 through 2011. After the final superconducting magnets were installed, the Tevatron set an energy record by accelerating a beam of protons to 512 GeV and then 800 GeV in 1984. The first proton-antiproton ($p\overline{p}$) collisions were recorded by the Collider Detector at Fermilab (CDF) in 1985 with a total center of mass energy of $\sqrt{s}$ = 1.6 TeV, increasing to $\sqrt{s}$ = 1.8 TeV a year later. After upgrades to the accelerator facilities, collisions at $\sqrt{s}$ = 1.96 TeV began in 2001 and continued through 2011. The Tevatron was the highest energy particle accelerator in the world before operation of the Large Hadron Collider (LHC) at CERN began in 2007.

What follows is a brief description of the accelerator chain that produced the high energy $p\overline{p}$ collisions analyzed here. We describe the proton source, antiproton source, Main Injector, and Tevatron. Further detail is provided in [37]. A diagram of the accelerator structure is shown in Fig. 2.1.

2.1.1 The Proton Source

The production of protons begins with a source that converts hydrogen gas into ionized hydrogen gas (H^-) inside a Cockroft-Walton generator. The dome of this generator, also

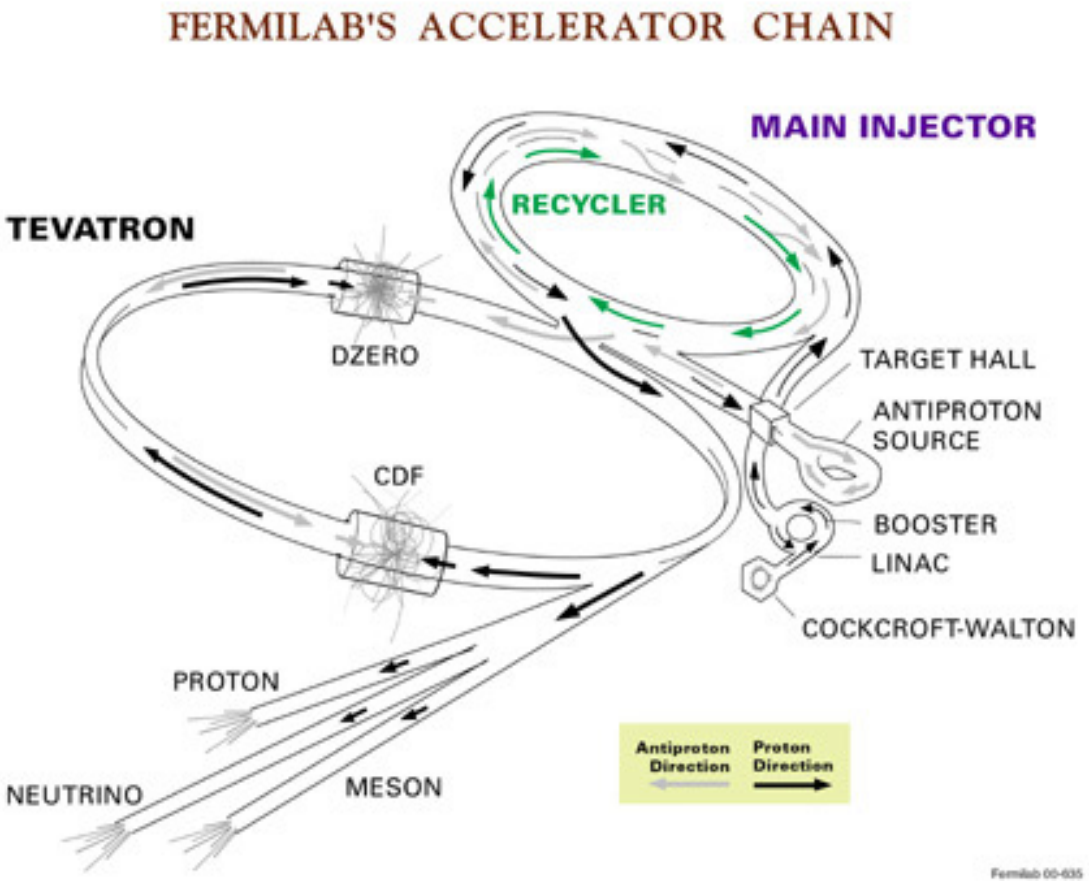

Figure 2.1: The chain of accelerators that produce the high energy $p\overline{p}$ collisions [38].

called the Pre-accelerator (or “Preacc”), is charged to a potential of -750 kV. The Preacc accelerates the ionized gas through a column between the dome and a grounded wall, giving the gas a final energy of 750 keV.

After the initial acceleration of ionized hydrogen, the ions travel to a linear accelerator, called the “Linac”. The Linac is composed of two main sections. The first uses a series of cavities equipped with drift tubes to shield particles from electric fields that would deaccelerate the ions. The second section is a side-coupled-cavity Linac, wherein each cavity is equipped with 16 smaller accelerating cells that are coupled together and powered by one RF source. The Linac also contains a number of magnets to focus the beam. The Linac increases the energy of the ionized hydrogen from 750 keV to 400 MeV.

When producing protons for the Tevatron, the beam from the Linac is sent through a carbon foil and into the “Booster”. The carbon foil strips the H^- ions of their electrons, leaving only protons. The Booster, a 75-m radius synchrotron, then accelerates the remaining protons to an energy of 8 GeV. Each of the elements of the proton source functions at 15 Hz, thus producing a beam of 8 GeV protons every 66 ms. Many protons from the Booster are used to create neutrinos for MiniBooNE and other experimental programs. Those used for high-energy collisions at CDF, however, are sent along to the Main Injector.

2.1.2 The Main Injector

The Main Injector is a large synchrotron that is about half the circumference of the Tevatron. Its eighteen accelerating cavities can take protons from the Booster, with energy 8 GeV, and accelerate them to either 120 or 150 GeV in as little as 2.2 s. When preparing for collisions in the Tevatron, the Main Injector accepts seven proton bunches from the Booster. It then accelerates these proton bunches to 150 GeV before combining them

into one larger bunch through a process called coalescing and sending them to the Tevatron for further acceleration. The Main Injector also accepts, accelerates, and injects into the Tevatron bunches of antiprotons that come from the Antiproton source.

2.1.3 The Antiproton Source

The creation of antiprotons for use in the Tevatron begins with 120 GeV protons, which come from the Main Injector and strike a stationary nickel alloy target. These collisions create a spray of secondary particles including antiprotons. Magnets extract antiprotons with an energy of 8 GeV from this particle spray, and these antiprotons are sent along a transfer line to the Debuncher.

The Debuncher is the first accelerator that makes up the antiproton source. It is a triangular-shaped synchrotron (with rounded edges) with a mean radius of 90 m. It captures the 8 GeV antiprotons which, coming from the target, have a large spread in their momentum. The Debuncher does not change the energy of the antiproton bunch; rather, the Debuncher stochastically cools the beam, using an electrical signal from the antiproton bunches circling the Debuncher to trigger a kick at another location in the accelerator. The cooling systems in the Debuncher reduce both the horizontal and vertical spread of the antiproton beam, as well as reduce the spread in the momentum of the antiprotons. From the Debuncher, the antiprotons enter the Accumulator, a similar synchrotron housed in the same tunnel as the Debuncher, which continues to cool and store the 8 GeV antiprotons.

The antiprotons are then sent to the Recycler. Originally intended to “recycle” antiprotons from the Tevatron, the Recylcer is used as a storage ring for antiprotons. The Recycler cools the antiproton beam further, using stochastic cooling as well as electron cooling. In the latter, a beam of electrons inserted with the antiprotons reduces the mo-

mentum spread of the antiproton beam through glancing collisions that transfer momentum from the antiprotons to the much lighter electrons. The Recycler, like the Debuncher and Accumulator, does not increase the energy of the antiprotons but keeps them at an energy of 8 GeV. When it is time to load antiprotons into the Tevatron, the antiprotons in the Recycler are first transferred to the Main Injector, then accelerated to 150 GeV, and finally sent along to the Tevatron.

2.1.4 The Tevatron

The Tevatron is the final machine in the accelerator chain for producing high-energy $p\bar{p}$ collisions at Fermilab. With a circumference of 6.3 km, it is the third largest synchrotron in the world and the largest in the Western Hemisphere. Its eight accelerating cavities take protons and antiprotons from the Main Injector and accelerate them from 150 GeV to an energy of 980 GeV, producing collisions with center-of-mass energy $\sqrt{s}$ = 1.96 TeV. The Tevatron uses a series of superconductor magnets to bend and focus the proton and antiproton beams as they travel around the accelerator. These magnets are made using niobium/titanium alloy wires that must be cryogenically cooled in order to be superconducting, allowing very large currents to pass through with minimal resistance. The magnets are cooled using liquid helium at a temperature of about 4 K.

The accelerator's RF frequency is 53.1 MHz, and so proton and antiproton bunches are collected in RF buckets with a temporal size of about 19 ns. Around the entire accelerator, there are 1113 evenly-spaced RF buckets. When colliding protons and antiprotons, each bunch is localized to a single bucket. Proton bunches are organized into 3 "trains" that contain 12 bunches each. Each bunch spaced 21 buckets (396 ns) apart, and between each train is an abort gap of 139 buckets (2.6 μs). Antiproton bunches are organized in the same way, moving around the ring in the opposite direction of the pro-

tons. With this bunch structure, $p\overline{p}$ collisions happen every 396 ns when the proton and antiproton trains intersect; on average, collisions happen every 590 ns, or at a rate of 1.7 MHz. Collisions occur at two interaction points around the ring: B0, the location of the CDF experiment, and D0, the location of the similarly-named experiment.

Once the protons and antiprotons have been accelerated, the Tevatron acts as a storage ring while collisions take place. The period for which a given bunch of protons and antiprotons is allowed to collide is called a "store". Typically, about 10^{13} protons and 0.3×10^{13} antiprotons are input into the Tevatron at the beginning of a store, which corresponds to an initial instantaneous luminosity of about 3×10^{32} cm^{-2}s^{-1}, or 30 nb^{-1}s^{-1}. Fig. 2.2 shows the initial instantaneous luminosity of stores during Run II of the Tevatron. The peak luminosities increased over time due to the efforts of Fermilab's accelerator experts. Over the lifetime of a single store, the instantaneous luminosity drops as particles are lost and the beams become less collimated. Once enough antiprotons for the next store have been produced, the colliding store is dropped and the remaining protons and antiprotons are sent into beam dumps by fast kicker magnets. Stores typically last $\sim$ 24 hours. Over the lifetime of the Tevatron, $p\overline{p}$ collisions with a total integrated luminosity of nearly 12 fb^{-1} were provided to the CDF and D0 experiments. The integrated luminosity collected over the decade the Tevatron ran during Run II is shown in Fig. 2.3.

2.2 The CDF II Detector

The CDF II detector [40] is a general-purpose particle detector located at one of the collision points of the Tevatron accelerator. The detector is roughly cylindrically symmetric around the beam line. Like most general-purpose detectors, CDF is made up of many sub-detectors spanning a variety of particle detection techniques that, when used together, can provide both accurate identification and measurement of the many particles

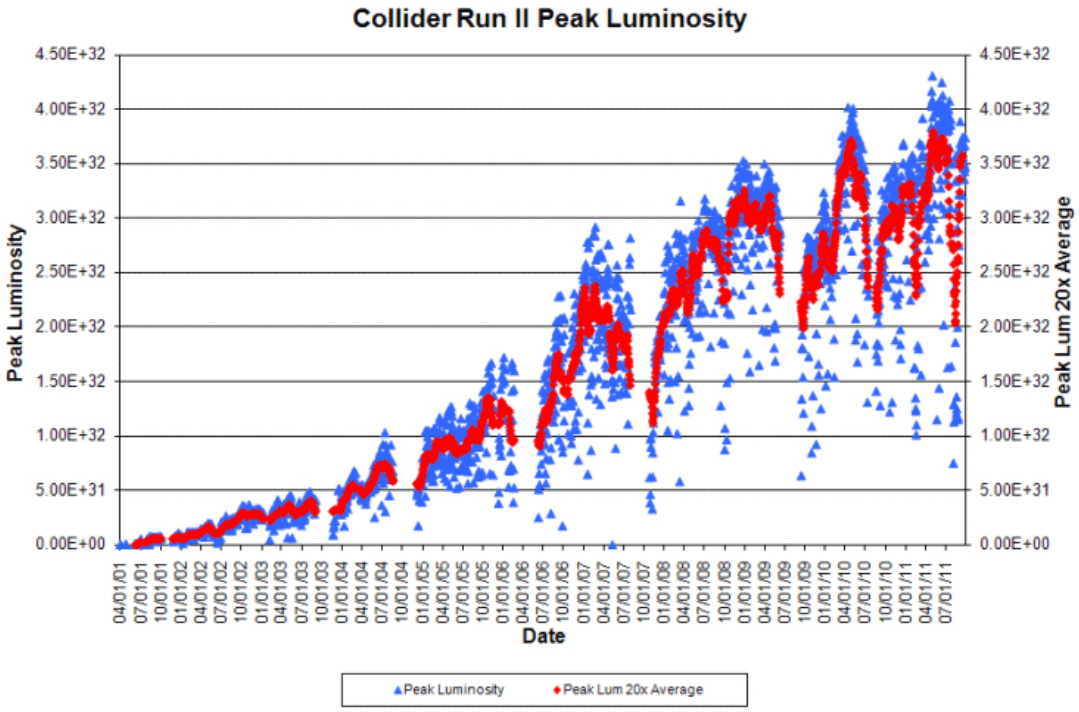

Figure 2.2: The peak instantaneous luminosity during each store (blue triangles) and a 20-store average (red diamonds) over time during Run II of the Tevatron [39]. The instantaneous luminosity is given in units of $cm^{-2}s^{-1}$. Periods of no data correspond to stoppages, when the Tevatron was not producing $p\overline{p}$ collisions.

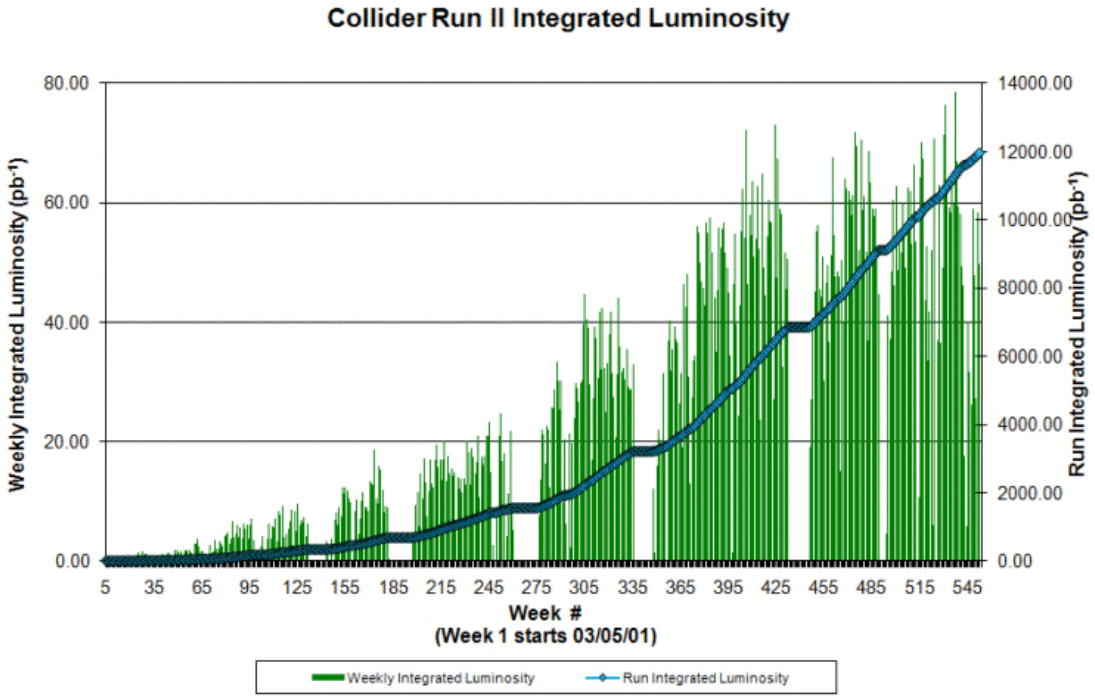

Figure 2.3: The integrated luminosity during each weak (green bars) and in total (cyan diamonds) during Run II of the Tevatron [39]. The integrated luminosity is given in units of inverse *picobarns*, where 1 $\text{pb}^{-1} = 10^{36}\ \text{cm}^{-2}$. Periods of no data correspond to stoppages, when the Tevatron was not producing $p\bar{p}$ collisions. In total, the Tevatron delivered $\sim 12\ \text{fb}^{-1}$ of $p\bar{p}$ collisions to the CDF and D0 experiments.

produced in $p\overline{p}$ collisions.

CDF is arranged in the following way. Charged particle tracking detectors are located outside the beamline, which contains the point of interaction for $p\overline{p}$ collisions. These tracking detectors reside within a superconducting solenoid that produces a 1.4 T magnetic field aligned coaxially with the proton and antiproton beams. This magnetic field causes charged particles to curve as they traverse the tracking systems. Around the outside of the solenoid, calorimeter modules measure the energies of charged and neutral particles. Finally, a series of drift chambers sit outside the calorimeter, and are used to detect muons as they typically leave little energy in, and are not stopped by, the calorimeters. An elevation view of the CDF detector is shown in Fig. 2.4.

Given the cylindrical geometry of the detector, the CDF experiment typically uses a cylindrical coordinate system to describe kinematic quantities. The z-axis of the detector is aligned with the beam line, with the origin of the coordinate system located at the detector's center and the incoming proton beam direction defined as the $+z$ direction. While not typically used to describe particles, r is the radial distance from the beam axis. The polar angle, θ, is measured from the origin with respect to the z-axis, and ϕ is defined as the azimuthal angle. Pseudorapidity is defined as $\eta \equiv -\ln\left(\tan\frac{\theta}{2}\right)$. The rectangular coordinates x and y point radially outward and vertically upward from the Tevatron ring, respectively, and together make up the "transverse" plane. Thus, a particle's transverse momentum, the component of momentum in this plane perpendicular to to the beam axis, is defined as $p_T = |\vec{p}|\sin\theta = p\sin\theta$. Similarly, we define the transverse energy of a electromagnetic or hadronic shower as $E_T = E\sin\theta$.

In the following sections we describe in greater detail each of the components of the CDF detector, as well as the trigger system, which is used to select the events of greatest interest to the experiment.

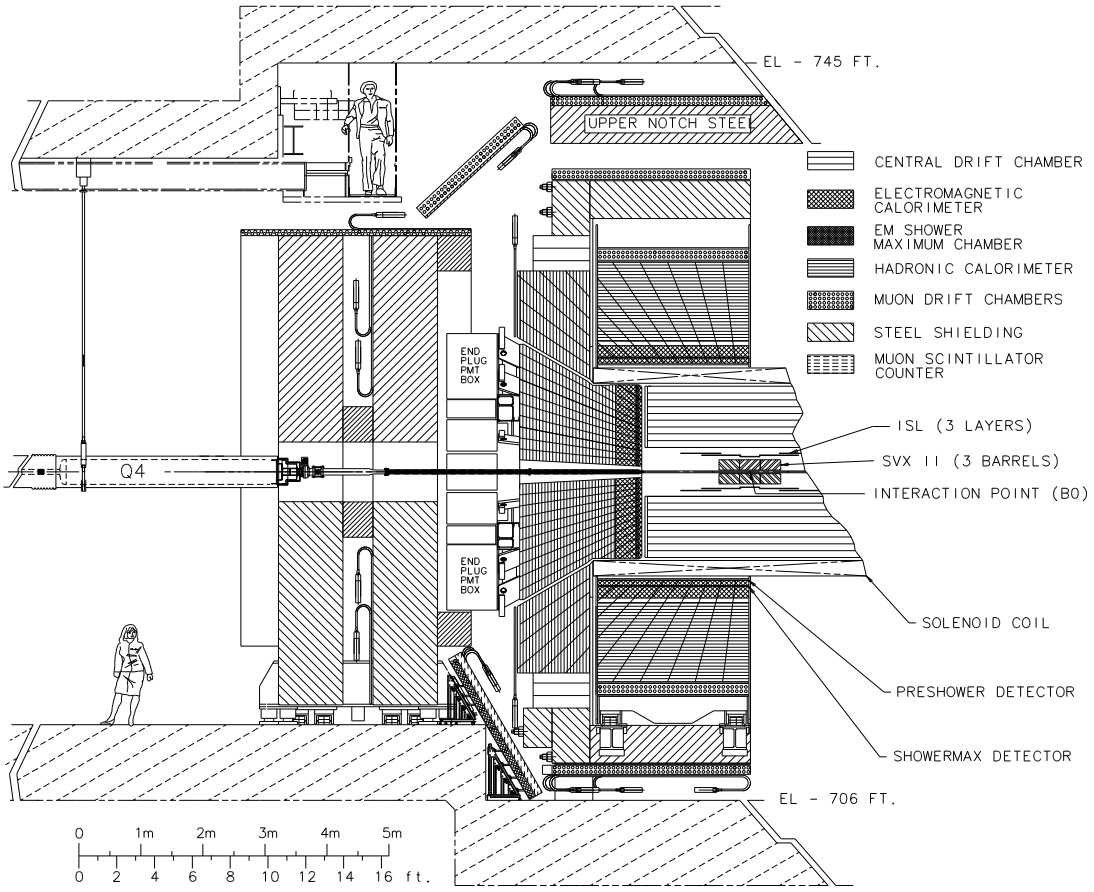

Figure 2.4: An elevation view of the CDF II detector. Various components of the detector are labeled and described in further detail in the text.

2.2.1 Tracking

The innermost tracking layers consist of three concentric silicon detectors [41]—Layer 00, SVX II, and Intermediate Silicon Layers (ISL)—ranging out to r = 30 cm and covering a pseudorapidity range of $|\eta| < 2$. The SVX II is a five-layer double-sided silicon microstrip detector in the region between 2.4 cm and 10.7 cm from the beam axis. It consists of three separate barrel modules that cover a length of 96 cm along the beam line, which is approximately 90% of the luminous interaction region. Three of the five layers combine an r-ϕ measurement on one side and a 90° stereo measurement on the other. The remaining two layers combine an r-ϕ measurement and a small-angle ($\pm$1.2°) stereo measurement. The typical hit resolution from the SVX is 11 μm. Layer 00 is single-sided silicon layer practically resting on the beam pipe, providing improved impact parameter resolution for reconstructed charged particle tracks. The ISL, another double-sided microstrip detector but with a larger pitch between strips than the SVX, covers radii between 19 and 30 cm from the beam line, and allows for the linking of tracks from the inner silicon detectors to the outer tracking chamber (described below).

Just outside the ISL sits the Central Outer Tracker (COT) [42], a 3.1 m long open cell drift chamber covering the region 0.40 m $< r <$ 1.32 m. It provides full coverage for charged particle tracks with $|\eta| < 1$. Sense wires are arranged in eight alternating axial (parallel to the z-axis) and $\pm$ 2° stereo (with respect to the z-axis) "superlayers." Each superlayer contains 12 wire layers, providing a total of 96 sense wire layers across the drift chamber. The chamber is filled with a 50:50 mixture of argon and ethane gas in which electrons have a drift time of about 200 ns, less than the 396 ns bunch-spacing of the proton and antiproton beams. The position resolution of single drift time measurements, measured using muons from $Z \to \mu^+\mu^-$ decays, is about 180 μm.

Two complementary algorithms associate segments lying on a common circle, and

the results are merged to form a final set of axial tracks segments. Tracks are reconstructed in three dimensions by associating track segments in stereo superlayers with the axial track segments. The efficiency of finding isolated high-momentum tracks is measured using electrons from $W^{\pm} \rightarrow e^{\pm}\nu$ decays identified in the central region ($|\eta| \leq 1.1$) using only calorimetric information from the electron shower and the missing transverse energy. The efficiency for finding these electron tracks is $99.93^{+0.07}_{-0.35}$%, and this is typical for other isolated high-momentum tracks from either electronic or muonic W and Z decays contained in the COT. The transverse momentum resolution of high-p_T tracks is $\delta p_T/p_T^2 \approx 0.1\%\ (\mathrm{GeV}/c)^{-1}$. Their track position resolution in the direction along the beam line at the origin is $\delta z \approx 0.5\,\mathrm{cm}$, and the resolution on the track impact parameter, the distance from the beam line to the track's closest approach in the transverse plane, is $\delta d_0 \approx 350\ \mu\mathrm{m}$.

To combine information from the silicon detectors to tracks reconstructed in the COT, a progressive "outside-in" tracking algorithm is used, in which COT tracks are extrapolated into the silicon detector, associated silicon hits are found, and the track is refit with the added information of the silicon measurements. The initial track parameters provide a width for a search road in a given layer. Then, for each candidate hit in that layer, the track is refit and used to define the search road into the next layer. This stepwise addition of precision SVX information progressively reduces the size of the search road, while also accounting for the additional uncertainty due to multiple scattering in each layer. The search uses the two best candidate hits in each layer to generate a small tree of final track candidates, from which the tracks with the best χ^2 are selected. The efficiency for associating at least three silicon hits with an isolated COT track is 91% $\pm$ 1%, and the extrapolated impact parameter resolution for high-momentum outside-in tracks is much smaller than for COT-only tracks: 30 μm, which includes the uncertainty in the beam position.

2.2.2 Calorimetry

Calorimeter modules sit outside the central tracking volume and solenoid, where they measure the energies of both charged and neutral particles. Modules consist of two main sections: inner electromagnetic (EM) layers consist of lead sheets interspersed with plastic scintillator; and, outer hadronic (HAD) layers consist of scintillators sandwiched between steel sheets. Electrons and photons, as they travel through the lead, undergo a series of bremsstrahlung and pair production, producing electromagnetic showers that are then measured by the scintillators. Electromagnetic showers are largely contained in the EM portions of the calorimeter. Hadrons also produce showers as they travel through the calorimeter, but these showers are typically much larger and wider, as they are created by inelastic scattering with nuclei in the absorber material.

The calorimeter is split between the central barrel ($|\eta| \leq 1.0$, with the electromagnetic and hadronic sections called the CEM and CHA, respectively) and forward end plug ($1.1 \leq |\eta| \leq 3.64$, similarly called the PEM and PHA) sections. An additional hadronic calorimeter (the end-wall hadronic calorimeter, or WHA) ensures complete coverage in the region $0.6 \leq |\eta| \leq 1.3$. The calorimeter modules are arranged in a tower based projective geometry. Individual towers in the central barrel subtend 0.1 in $|\eta|$ and 15° in ϕ. The sizes of the towers in the end plug calorimeter vary with $|\eta|$: subtending 0.1 in $|\eta|$ and 7.5° in ϕ at $|\eta| = 1.1$, and 0.5 in $|\eta|$ and 15° in ϕ at $|\eta| = 3.6$.

The energy resolution of the calorimeter towers was measured using test beam data. For electrons, the energy resolution is $14\%/\sqrt{E_T}$ in the CEM and $16\%/\sqrt{E} \oplus 1\%$ in the PEM, where the energies are in units of GeV. The single-particle energy resolution in the hadronic calorimeters, measured using pions, is $75\%/\sqrt{E}$ in the CHA, $80\%/\sqrt{E}$ in the WHA, and $80\%/\sqrt{E} \oplus 5\%$ in the PHA. Further studies have been done to determine and validate the energy scale of hadronic jets, and are explained in greater detail in Chap. 5.

In addition to the scintillators that measure particle energies, both the CEM and PEM have electromagnetic shower maximum detectors (the CES and PES, respectively) imbedded at the point where an electromagnetic shower reaches its maximum intensity. The CES is a proportional chamber with wires in the r-ϕ view and cathode strips in z. It accurately measures the position of the shower in each dimension, and this position may then be matched to a track in the COT if the originating particle was an electron. The position resolution in r-ϕ is about 0.2 cm. The PES contains two layers of scintillator strips, oriented 45° with respect to each other. Also, in front of both the CEM and PEM there are pre-radiator detectors (the CPR and PPR) that provide additional information before showering starts in the calorimeter.

2.2.3 Muon detectors

Because they have a higher mass, muons do not radiate much at momenta $< 10^2$ GeV/c, and therefore they do not produce showers in the electromagnetic calorimeter like electrons. Instead, they pass through the calorimeters largely undetected. So, outside the calorimeters, a collection of drift chambers are used to detect muons. Steel absorbers reduce the likelihood for a hadronic jet to punch through the calorimeter and into the muon detectors, while not stopping muons with $p_T > 3.0$ GeV/c. Muons may then be identified by their tracks in the COT matched to "stubs" found in the muon chambers.

There are three major muon detectors used in this analysis. The central muon detector (CMU) is located directly outside of the central calorimeter. It is a four-layer stack of planar drift chambers. Position measurements in r-ϕ are made by converting the signal arrival time into a drift distance. The maximum drift time in a CMU cell is 800 ns. This is longer than the bunch spacing (396 ns), but any ambiguity to the proper $p\overline{p}$ interaction can be made by identifying an associated COT track produced by the muon. The CMU covers a

range of $|\eta| < 0.6$.

The central muon upgrade detector (CMP) and central muon extension detector (CMX) are also both wire drift chambers, though wider than the CMU detector (and thus have a longer drift time). The CMP is an additional four-layer stack of drift chambers, arranged in a box-like structure around the CMU behind 60 cm of steel absorbers. It provides greater background rejection and some additional coverage for $|\eta| < 0.6$. The CMX detectors consists of eight layers of drift cells in a conical arch-like arrangement around the detector, thus giving some position measurement in the z direction from overlapping cells with different stereo angles with respect to the beam line. It extends the muon coverage up to $|\eta| < 1.0$.

2.2.4 Luminosity Measurement

The small-angle Cherenkov luminosity counters (CLC) detector measures the instantaneous luminosity and integrated luminosity of our data sample. It consists of two modules of isobutane gas Cherenkov counters pointing toward the interaction region, located around the beam pipe at each end of the detector, covering $3.6 < |\eta| < 4.6$. The CLC directly measures the rate of inelastic $p\overline{p}$ events, $R_{p\overline{p}}$. This may be translated into a luminosity knowing the total inelastic $p\overline{p}$ cross section, taken to be $\sigma_{p\overline{p}} = 60.7 \pm 2.4$ mb, and the acceptance of the CLC, determined to be $A_{CLC} = 60.2 \pm 2.6\%$:

$$\mathcal{L} = \frac{R_{p\overline{p}}}{A_{CLC} \times \sigma_{p\overline{p}}} .$$

Further details of the luminosity measurement are described in [40].

2.2.5 Trigger System

The rate at which data from the detector may be read out and recorded for use in later analyses is limited: $\sim$ 100 Hz. Since collisions happen at a rate of about 2.5 MHz (though 1.7 MHz on average, due to the beam structure), a trigger system selects only a portion of the $p\bar{p}$ collisions to be stored to tape. This system must operate in real-time, and so be fast enough to handle the very high rate of collisions, but also be sophisticated enough to identify the events that appear most deserving for future analysis.

CDF employs a three-level trigger system utilizing a combination of dedicated hardware and specialized software running on commercial computational processors. An overview of the data flow is shown in Fig. 2.5. Synchronous with each beam crossing, signals from the detectors are digitized and saved in a storage pipeline (with the exception of the silicon detectors, which use an analog pipeline). Meanwhile, the data needed to make the first level (L1) trigger decision is sent to three sets of custom hardware used to identify objects of interest in the event: calorimeter objects (electrons, photons, and hadronic jets), stubs in the muon detectors, and tracks in the COT. A global trigger decision is then made, based on the number of physics objects in the event and/or global calorimeter energy quantities. Upon a L1 accept, the silicon data is digitized, and all event data is removed from the L1 pipeline and stored in one of four event buffers in the front-end electronics of the detector components. The L1 accept rate is on the order of 20 kHz.

Unlike L1, the second level trigger (L2) operates asynchronously with the beam crossing. After an L1 accept, data from the silicon detectors is sent to the silicon vertex trigger (SVT) [43, 44], which uses dedicated hardware to perform fast and accurate track-fitting using the silicon data. Data from other detector components is collected and formatted on PULSAR boards [45], which in turn send the data to a PC that runs trigger algorithms

(like basic jet clustering, momentum imbalance calculations) with higher precision than can be done at L1. The global latency for L2 is $\sim 40\mu$s, and it has a typical accept rate of $\sim$ 0.5-1 kHz. The possibility of deadtime arises in the transition from L1 to L2: if the L1 accept rate is too high or the processing time at L2 is too long, the four front-end buffers may fill up. CDF has designed its trigger system to operate with a deadtime of $\lesssim$ 5%.

Upon a L2 accept, the data from the front-end buffers on the detector are read out and sent to the level 3 trigger (L3). L3 is a processing farm of nearly 300 commercial dual processor computers. Each runs event reconstruction code of near-offline quality and imposes a loose set of cuts on reconstructed objects, and can process an event in about 1 s. With the full detector information available to it, and a longer latency at which it may operate, the L3 algorithms can apply even more precise cuts to remove events faking the signals associated with the objects we desire to store. The typical L3 accept rate is $\sim$ 100-150 Hz. After an L3 accept, the data is sent to storage areas, where it is then analyzed offline.

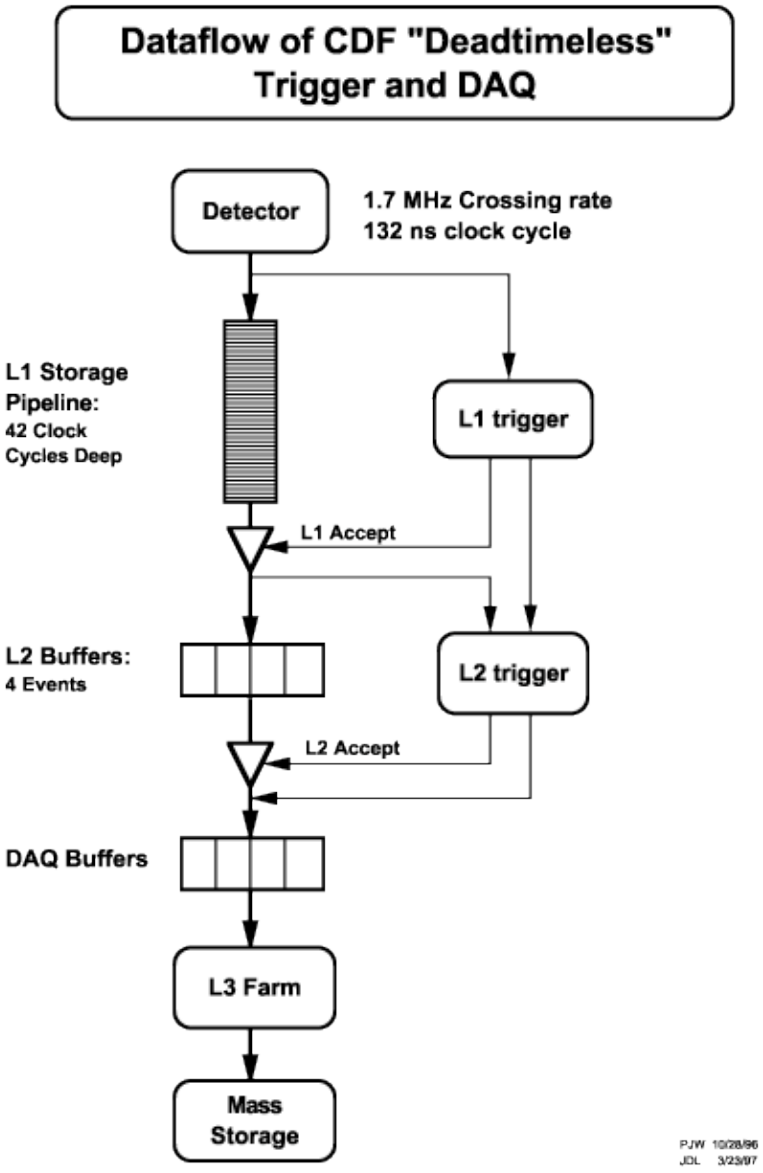

Figure 2.5: The data flow in CDF's three-level realtime trigger system, which reduces the 1.7 MHz average event rate to about 100 Hz. The data is then be stored for further offline analysis.

CHAPTER 3

EVENT RECONSTRUCTION

After events are recorded, they are analyzed offline to find physical processes of interest. In this analysis, we search for ZW and ZZ production, where a Z boson decays to an e^+e^- or $\mu^+\mu^-$ lepton pair, and the other boson decays to a quark-antiquark pair: either $Z \to q\bar{q}$, or $W^{\pm} \to q\bar{q}'$. In order to identify these events, we need to reconstruct each of the decay products from diboson production. This chapter contains a summary of the ways in which we reconstruct these, and other, objects based on their signatures in the CDF detector.

3.1 Electrons

High-energy electrons and positrons (hereafter, we will not consider the matter/antimatter distinction and will call all such objects "electrons"), as they pass through the detector, should leave the following signatures in the CDF detector: a charged particle track in the silicon detectors and the COT, an electromagnetic shower in the EM calorimeter, and very little energy deposited in the hadronic calorimeter. The selection of our electrons is based on identifying each of these signatures in the event.

3.1.1 Central Electrons

For central electrons ($|\eta| < 1$), we search for tracks found in the COT. The track quality can be quantified by counting the number of hits associated with the charged particle track.

We typically count the number of individual hits seen in each superlayer: if seven or more out of the twelve wires receive a significant signal, we consider that superlayer segment as part of the track. We can then place cuts on the number of axial and stereo segments contained in the track: typically 3 axial segments and 2 stereo segments gives a track of good quality. With a reconstructed track, we also extrapolate and find the z coordinate of where the track intersects with the beamline in the r-z plane. We can require this quantity, the track z_0, to be consistent with the location of the luminous region: within 60 cm of the detector's center.

Central electrons should also leave energy in the CEM. Electron clusters are formed by searching for calorimeter towers with EM $E_T > 2$ GeV and a COT track pointing to the same tower. One additional tower may be included in the electron cluster if is on the same ϕ wedge and adjacent to the seed tower in η. Since electrons should deposit little energy in the CHA, the electron cluster is required to have a total hadronic energy that is less than 12.5% of the electromagnetic energy.

Once we have identified an electromagnetic cluster, quality requirements can be made on variables that distinguish electrons from photons or jets. Typically, a stricter requirement on the ratio of hadronic to electromagnetic energy is made: $E_{HAD}/E_{EM} < 0.055 + 0.00045 \cdot E$ [GeV]. Also, we can require the electron to be "isolated" in the calorimeter: that any other energy in a cone of $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \leq 0.4$ is less than 10% of the E_T of the electron.

For electrons with well-reconstructed tracks, we can examine the ratio of the electromagnetic energy to the momentum of the COT track: E/p. For relativistic electrons, this quantity should be 1, while photons without a COT track associated to them (or an essentially random COT track associated with them) will typically have much higher E/p. However, high-energy electrons may emit a collinear photon in the tracking volume that decreases the track momentum. For our highest-quality electron selection, we place a cut

on E/p at less than 2.5 plus some small fraction of the track p_T.

Finally, we can place cuts based on properties of the electromagnetic shower. The lateral shower profile can be compared with test-beam data, and a varaible L_{shr} [46] can be used to check consistency with the expected behavior of electrons. Also, with a reconstructed COT track, we can match properties of the track to the location of the center of shower maximum in the CES. We can measure the distance in the r-ϕ plane between the extrapolated COT track location and the cluster in the CES, $Q \cdot \Delta x_{CES}$ (it is multiplied by the charge of the track to account for asymmetries in the tail of the track from bremsstrahlung), as well as the distance in the r-z plane, Δz_{CES}.

We define three different types of central electron: Tight Central Electron (TCE), Loose Central Electron (LCE), and Central Trackless Electron objects (CTE). The requirements for an electron to be classified as one of these types are given in Tab. 3.1. TCE objects are our highest quality central electrons. The LCE selection removes some of the quality requirements on the shower profile and E/p. A CTE object is, basically, an electromagnetic cluster with a track of minimal quality pointing to it (so, despite their name, they are not truly trackless). These electron classes are forced to be mutually exclusive.

3.1.2 Forward (Plug) Electrons

For forward regions of the detector ($|\eta| > 1$), the COT does not provide full coverage, and so we cannot solely rely on COT tracks for electrons in this region. We can, instead, search for hits in the silicon detectors, which owing to their position close to the beam line provide much greater η, coverage. These "PHOENIX" tracks can be matched to clusters in the PEM, and thus improve electron identification in forward regions of the detector. We only consider electrons with $|\eta| \leq 2.8$, as beyond this region the calorimeter is not modeled as well, and there are an increased number of hadronic jets that can fake our

	TCE	LCE	CTE
$E_{\mathrm{HAD}}/E_{\mathrm{EM}}$	$< 0.055 + 0.00045 \times E$ [GeV]	$< 0.055 + 0.00045 \times E$ [GeV]	$< 0.055 + 0.00045 \times E$ [GeV]
Iso. Ratio ($\Delta R \leq 0.4$)	< 0.1	< 0.1	< 0.1
Found COT Track?	Yes	Yes	Yes
Track $\lvert z_0 \rvert$	< 60 cm	< 60 cm	< 60 cm
Track p_T	> 10 GeV/c	> 10 GeV/c	
N Axial Seg.	≥ 3	≥ 3	
N Stereo Seg.	≥ 2	≥ 2	
E/p	$< 2.5 + 0.015 \times p_T$ [GeV/c]		
L_{shr}	< 0.2		
$Q \cdot \Delta x_{CES}$	> -3.0 cm, < 1.5 cm		
$\lvert \Delta z_{CES} \rvert$	< 3.0 cm		

Table 3.1: Summary of the cuts on electromagnetic clusters used to define our central electron types in this analysis.

electron selection.

Similar to how we form electron clusters in the CEM, we also form such clusters in the PEM. Here, we allow the clusters to contain a seed tower plus two additional towers, adjacent in η or ϕ to the seed. As in the central electron selection, electron clusters in the PEM are required to have a total hadronic energy that is less than 0.125 times the electromagnetic energy. We can place cuts similar to the central electron selections: we typically require $E_{HAD}/E_{EM} < 0.05$ and that the cluster be isolated. Additionally, we can look at a 3×3 array of towers around the seed tower, and determine if the distribution of energies in this array is consistent with the what we expect for electrons based on test-beam data via a χ^2_{PEM} comparison. We also may consider showering information from the PES detectors, looking at how well the shower profile matches the energy distribution in the PEM towers. This can be checked in both sets of PES strips (labeled *u* and *v*).

We define two different types of forward electrons: those with silicon hits associated with a high-energy PEM cluster and those without such a track. The former we call PHOENIX (or PHX) electrons, while the latter we call simply plug electromagnetic (PEM) objects. A summary of the cuts and requirements for these electron types is shown in Tab. 3.2.

	PHX	PEM
$\|\eta\|$ range	$1.13 < \|\eta\| < 2.4$	$1.13 < \|\eta\| < 2.8$
$E_{\mathrm{HAD}}/E_{\mathrm{EM}}$	< 0.05	< 0.05
Iso. Ratio ($\Delta R \leq 0.4$)	< 0.1	< 0.1
χ^2_{PEM}	< 25	< 25
PES(u) 5×9 Shower Profile Ratio	< 0.65	< 0.65
PES(v) 5×9 Shower Profile Ratio	< 0.65	< 0.65
ΔR btw. PES and PEM Cluster	< 3.0 cm	< 3.0 cm
N_{hits} in silicon	≥ 3	
Track $\|z_0\|$	< 60 cm	

Table 3.2: Summary of the cuts on electromagnetic clusters used to define our forward (or plug) electron types in this analysis.

3.2 Muons

High-energy muons and antimuons (again, we will hereafter call all such objects "muons") should leave the following signature as they pass through the detector: a charged particle track in the silicon and COT, very little energy in the electromagnetic and hadronic calorimeters, and hits in the muon detectors sitting outside the calorimeter if the muons are traveling towards those detectors. Thus, an essential part to muon reconstruction is finding a high-quality track that points to little energy in the calorimeters. The presence of hits in the muon detectors associated with that track provide higher assurances that this track is from a muon.

The track requirements for central muons mirror many of the requirements of central electrons: we require a number of axial and stereo COT segments to contain the track and that the track z_0 lie within the luminous region of the beam line. However, we require additional constraints on central muon tracks. We can reconstruct the impact parameter, d_0 of that track with respect to the beam line. High-p_T muons produced in the primary $p\bar{p}$ interaction should have $d_0 \approx 0$, while muons from K or π meson decays or from cosmic rays may have much higher impact parameters. The cut we place on the impact parameter depends on whether or not silicon hits are attached to the COT track; if they are, the impact parameter resolution is much improved. Additionally, for muons not from the primary $p\bar{p}$ interaction, the track fit will be of worse quality—we can place a cut on the χ^2 of the fit to the track to reduce contributions from these muons. We also place caps on the energy in the electromagnetic and hadronic calorimeter near the reconstructed track.

We define muon types largely by the muon detector (CMU, CMP, or CMX), if any, where they leave hits. Because the η coverage of the CMU and CMP detectors overlap, we also define a high-quality class of muons labeled "CMUP". We match stubs to tracks in the r-ϕ plane, and look at the distance between stub hits and the track: ΔX. The matching is

required to be within about 7 cm. Additionally, for muons matched to hits on the CMX, we require the track pass through all eight COT superlayers by placing a requirement on the "exit radius" of the COT track, defined as

$$\rho_{COT} \equiv (z_{COT} - z_0) \cdot \tan\theta \ ,$$

where z_{COT} is the distance from the edge of the COT to the center of the detector: +155 cm for $\eta > 0$, and -155 cm for $\eta < 0$. A requirement that $\rho_{COT} < 140$ cm ensures that the track passes through a region that best matches the requirements of some of the high-p_T muon triggers.

We also look for muons that do not leave hits in the muon detectors, increasing our acceptance beyond the limited coverage of the muon detectors. For these "stubless" muons, we require there be a small amount of energy in the calorimeter (100 MeV) as muons should leave a small amount of energy as they pass through. For the forward regions of the detector, we require stubless muons leave well reconstructed track hits in 60% of the COT layers that the track should, based on its η. pass through. We additionally place caps on the electromagnetic and hadronic energies in the calorimeter, to reduce the number of jets that may fake these cuts. These central and forward stubless muons are called "CMIOCES" and "CMIOPES" muons, respectively.

Finally, we accept a class of objects that are simply well-reconstructed tracks that point to cracks in the calorimeter. These cracks may be between calorimeter tower wedges, between the two halves of the central calorimeter (at η = 0), or between the central and plug calorimeters. The tracks are simply required to not point to active regions in the CEM or PEM calorimeter, and that they, like electrons, pass calorimeter isolation requirements. These "Crack Track" (CrkTrk) objects may be either electrons or muons.

Tab. 3.3 contains a summary of the common requirements on our muon/CrkTrk ob-

jects, Tab. 3.4 summarizes the additional individual requirements on each of the stubbed muon types, and Tab. 3.5 the requirements on the stubless central muons and CrkTrk objects. Tab. 3.6 summarizes the requirements on stubless forward muons.

Common Central Muon/CrkTrk Cuts
N Axial COT Seg. ≥ 3
N Stereo COT Seg. ≥ 2
Corrected $d_0 < 0.02$ cm (< 0.2 cm if no silicon hits attached)
Track $\lvert z_0 \rvert < 60$ cm
χ^2/ n.d.o.f. $< 3.0^*$
$E_{\mathrm{EM}} < 2.0$ GeV + $0.0115 \times (p$ [GeV/c] $-100 GeV$/c$)^\dagger$
$E_{\mathrm{HAD}} < 6.0$ GeV + $0.028 \times (p$ [GeV/c] $-100 GeV$/c$)^\dagger$

Table 3.3: Summary of the common requirements for central muons and CrkTrk objects. Additional requirements for each of the muon objects are listed in Tab. 3.4 and Tab. 3.5.

*This requirement is loosened to < 4.0 for earlier runs.

†The additional leeway on the energy limits in the calorimeter applies only to tracks with $p > 100$ GeV/c.

3.3 Quarks and Gluons: Jets

Quarks and gluons (or "partons") do not appear as individual particles when they are produced in particle colliders due to the confinement properties of the strong force. Instead, single quarks and gluons will form showers of color-neutral mesons and baryons. This process, called hadronization, leads to the formation of "jets" of particles clustered together. This cluster of particles may leave a number of tracks in the tracking chamber, and then leave large clusters of energy in both the electromagnetic and hadronic portions of calorimeter towers. At CDF, we must identify the towers that form these jets, measure their energy properly, and then correct the energy of the jet to the match the energy of the originating parton. The jet clustering and energy corrections are described in this section and Chap. 5, respectively, and are described in further detail in [47] and [48].

Calorimeter towers are typically clustered to form jets using the JETCLU fixed-size cone

	CMUP	CMU	CMP	CMX
ΔX_{CMU}	< 7.0 cm	< 7.0 cm		
ΔX_{CMP}	$< max(6.0\text{ cm}, 150/p_T\text{ [GeV}/c\text{]})$		$< max(6.0\text{ cm}, 150/p_T\text{ [GeV}/c\text{]})$	
ΔX_{CMX}				$< max(6.0\text{ cm}, 125/p_T\text{ [GeV}/c\text{]})$
ρ_{COT}				< 140 cm

Table 3.4: Summary of the individual requirements for central stubbed muons. All muons also satisfy the requirements in Tab. 3.3.

	CMIOCES	CrkTrk
χ^2/ n.d.o.f	< 3.0	< 3.0
Track Fiducial to CEM?	Yes	No
Track Fiducial to PEM?	No	No
$E_{\text{EM+HAD}}$	> 100 MeV	> 100 MeV
Iso. Ratio ($\Delta R \leq 0.4$)		< 0.1

Table 3.5: Summary of the individual requirements for stubless central muons and CrkTrk objects. Each muon type also satisfies the requirements in Tab. 3.3.

CMIOPES Definition
Corrected $d_0 < 0.02$ cm (< 0.2 cm if no silicon hits attached)
Track $\lvert z_0 \rvert < 60$ cm
COT Hit Fraction > 0.6
$E_{\text{EM}} < 2.0$ GeV + $0.0115 \times (p$ [GeV/c] $- 100 GeV$/c)*
$E_{\text{HAD}} < 6.0$ GeV + $0.028 \times (p$ [GeV/c] $- 100 GeV$/c)*
$E_{\text{EM+HAD}} > 100$ MeV
Track Fiducial to PEM

Table 3.6: Summary of requirements for tracks to pass our stubless forward requirements (CMIOPES). (*) The additional leeway on the energy limits in the calorimeter applies only to tracks with $p > 100$ GeV/c.

*The additional leeway on the energy limits in the calorimeter applies only to tracks with $p > 100$ GeV/c.

algorithm [47]. For each event, all towers with $E_T = E \sin\theta = (E_{EM}+E_{had}) \sin\theta > 1$ GeV are collected in a seed tower list, sorted by decreasing tower E_T. Jets are formed by adding to the seed tower all towers within a given ΔR (the cone size) and with $E_T > 1$ GeV, giving an initial list of clusters. For each cluster, an ${E_T}^{\text{jet}}$, ${\eta}^{\text{jet}}$, and ${\phi}^{\text{jet}}$ are defined as follows:

$$ {E_T}^{\text{jet}} = \sum_{i=0}^{N_{tow}} E_{T,i} \ ; $$

$$ {\eta}^{\text{jet}} = \sum_{i=0}^{N_{tow}} \frac{E_{T,i} \cdot \eta_i}{{E_T}^{\text{jet}}} \ ; \text{and,} $$

$$ {\phi}^{\text{jet}} = \sum_{i=0}^{N_{tow}} \frac{E_{T,i} \cdot \phi_i}{{E_T}^{\text{jet}}} \ . $$

Here, N_{tow} is the number of towers associated with each cluster, and $E_{T,i}$, η_i, and ϕ_i are the transverse energy and position coordinates of tower i. If any jets overlap, they are merged if their overlapping towers make up more than 50% of each jet, otherwise each tower is assigned to the nearest jet. Once an initial list of jets is formed, the jet position coordinates serve as new seeds, and a new list of towers within the cone size is formed for each jet. The process of clustering and determining new jets is repeated until the towers in each cluster is stable. The final jet variables are then defined as follows:

$$ {E}^{\text{jet}} = \sum_{i=0}^{N_{tow}} E_i \ ; $$

$$ {p_x}^{\text{jet}} = \sum_{i=0}^{N_{tow}} E_i \sin\theta_i \cos\phi_i \ ; $$

$$ {p_y}^{\text{jet}} = \sum_{i=0}^{N_{tow}} E_i \sin\theta_i \sin\phi_i \ ; $$

$$p_z^{\text{jet}} = \sum_{i=0}^{N_{tow}} E_i \cos\theta_i \ ;$$

$$p_T^{\text{jet}} = \sqrt{(p_x^{\text{jet}})^2 + (p_y^{\text{jet}})^2} \ ;$$

$$\phi^{\text{jet}} = \tan \frac{p_y^{\text{jet}}}{p_x^{\text{jet}}} \ ;$$

$$\sin\theta^{\text{jet}} = \frac{p_y^{\text{jet}}}{\sqrt{(p_x^{\text{jet}})^2 + (p_y^{\text{jet}})^2 + (p_z^{\text{jet}})^2}} \ ; \text{ and,}$$

$$E_T^{\text{jet}} = E^{\text{jet}} \sin\theta^{\text{jet}} \ .$$

After the formation of jets using the clustering algorithm, a number of energy corrections, described in Chap. 5, are applied to jets with $E_T > 8$ GeV. The energy before corrections is called the "raw" energy, while the energy after corrections is called the "corrected" energy. Most of the jets used in this $ZW/ZZ \to \ell^+\ell^- + jj$ analysis have a cone size of $R = 0.4$, and have corrected $E_T > 25$ GeV and $|\eta| < 2.0$. Some portions of the analysis look at jets down to 15 GeV in transverse energy and $|\eta|$ up to 2.4, but jets beyond these boundaries are not as well modeled in our Monte Carlo simulations.

3.4 Neutrinos: Missing Transverse Energy

Neutrinos are not a part of our main analysis selection, but we do use $W \to \ell\nu$ and $t\bar{t} \to W^+bW^-\bar{b} \to b\bar{b}\ell\nu q\bar{q}'$ selections as calibration regions for some of the new tools developed for our analysis. Neutrinos interact with matter, but they do so at a rate so low that we may assume neutrinos escape without depositing any energy in our detector. In order to try to select events with a neutrino, we make use of the fact that the incoming protons and antiprotons have only momentum in the z direction. While we do not know the total p_z of the partons that take part in the primary interaction, we do know that the p_T of these partons is approximately zero. Thus, momentum in the transverse plane should

be conserved. A significant imbalance in the total transverse momentum would indicate that a particle produced in the collision escaped detection.

We define "missing energy" variables by taking the negative of the sum over all calorimeter towers (i) in the event:

$$\not\!\!E_x \equiv -\sum_i E_{T,i} \cos\phi_i$$

$$\not\!\!E_y \equiv -\sum_i E_{T,i} \sin\phi_i \ .$$

With these quantities we may define a missing transverse energy vector, $\vec{\not\!\!E}_T \equiv \not\!\!E_x\hat{\mathbf{i}} + \not\!\!E_y\hat{\mathbf{j}}$, whose magnitude, $\not\!\!E_T$, we call the missing transverse energy (or "MET"). Due to jet energy corrections, we must also "correct" this quantity to account for these corrections—the corrections are made up to the absolute jet energy scale (see Chap. 5, especially Sec. 5.4). We also correct the $\not\!\!E_T$ to account for identified muons in the event, as muons carry momentum that is largely unseen by the calorimeter. We associate this final, corrected missing energy with the p_T of an escaping particle, like a neutrino.

However, $\not\!\!E_T$ can also be created in events with multiple jets where one or more jet energies are not accurately measured, as this will create a "fake" momentum imbalance. While it is unlikely for such an event to yield large $\not\!\!E_T$—enough to mimic a neutrino or multiple neutrinos—the production cross section for multijet production is very high, and thus can become a significant backgrounds to events with "real" $\not\!\!E_T$. Thus, we may also consider the $\not\!\!E_T$-significance [31], a dimensionless quantity that characterizes how significant the reconstructed $\not\!\!E_T$ is compared to typical $\not\!\!E_T$ contributions from the poor energy resolution of jets, the presence of soft, unclustered particles, and to variations in event topologies.

CHAPTER 4

EVENT SELECTION AND MODELING

With the Tevatron having finished operations in September of 2011, we can perform this search for ZW/ZZ production in the charged leptons + jets channel using the full collection of $p\overline{p}$ collisions taken by the CDF II detector. Because we depend on the identification of both electrons and muons in our analysis, and we rely on the silicon detectors to improve our ability to identify b jets (see Chap. 6), we require the data to come from periods of operation when the calorimeter, muon detectors, and silicon detectors were all functioning properly. The integrated luminosity of this final dataset is 8.9 fb^{-1}.

4.1 Trigger Selection Requirements

As discussed in Sec. 2.2.5, it is impossible to permanently record every $p\overline{p}$ collision that takes place in CDF. Due to their distinct signatures in the detector, high-p_T leptons are often used to identify physical processes of interest and separate them from the high multijet backgrounds at hadron colliders. High-p_T leptons are used in many analyses at CDF: in studies of the top quark, the W and Z bosons, and in searches for the Higgs boson and physics beyond the Standard Model. Thus, triggering on high-p_T electrons and muons is a priority for the experiment.

We use data collected on a suite of high-p_T lepton triggers. The triggers with the largest contributions are:

- a central electron trigger that requires at least one electromagnetic cluster in the CEM matched to a high-p_T (> 18 GeV/c) COT track;

- a “Z No-Track” trigger that requires at least two electromagnetic clusters (in either the CEM or PEM) whose energies are consistent with being from a Z boson decay;
- and, central muon triggers that require at least one high-p_T (> 18 GeV/c) COT track matched to hits in the CMU and CMP or CMX detectors.

Additional triggers that accept other leptons (like muons which only leave hits in the CMU or CMP detectors) or leptons in association with other objects (like a lepton with a jet) also contribute to our sample. Some triggers in this selection are also pre-scaled, where only a fraction of the total number of events passing the trigger requirements are written to tape. This acceptance across a number of different triggers leads to an increase in the overall acceptance for our diboson signal.

4.2 Selection Requirements

The Feynman diagrams for ZW/ZZ production and decay to the $\ell^+\ell^-$ + jj final state are shown in Fig. 1.5. Our final state contains two oppositely-charged high-p_T leptons from the decay of a Z boson, and two (quark) jets from the decay of a W or Z boson. We include both W and Z boson decays because the resolution of the calorimeter does not allow us to distinguish between hadronic decays of the W boson ($m_W = 80.4$ GeV/c^2) and the Z boson ($m_Z = 91.2$ GeV/c^2).

The details of the event selection are summarized in Tab. 4.1. We demand both leptons have $p_T > 20$ GeV/c, so as to be in a kinematic region where the triggers are highly efficient. We require there to be a reconstructed interaction vertex in the luminous region of the detector, and that the individual leptons’ reconstructed crossing point with the beamline, z_0, be within 5 cm of each other for leptons with well reconstructed tracks: central electrons and all muons. Due to the poorer tracking information on forward electrons, we remove this requirement on electron pairs containing a PHX or PEM

electron. Similarly, we require lepton pairs with well-reconstructed tracks to have opposite charge. Most importantly, leptons are required to have a reconstructed dilepton mass in a 30 GeV/c^2window around the Z boson mass: 76 GeV/$c^2 \leq M_{\ell\ell} \leq$ 106 GeV/c^2.

Z + jets Selection
$N_{vert} > 0$
$\lvert z_0 \rvert <$ 60 cm
Lepton pair from Tab. 4.3
Δz_0 between leptons $<$ 5 cm*
Leptons oppositely charged*
1st/2nd lepton $p_T >$ 20 GeV/c
76 GeV/$c^2 < M_{\ell\ell} <$ 106 GeV/c^2
$\not\!\!E_T <$ 20 GeV
1st/2nd jet $E_T >$ 25 GeV
1st/2nd jet $\lvert\eta\rvert <$ 2.0
ΔR between jets and any lepton $>$ 0.4
ΔR between jets $>$ 0.7

Table 4.1: Summary of event selection requirements in the analysis.

*Only applied to central electron and muon pairs.

For the selection of the hadronically decaying W or Z boson, we require there be at least two jets with $E_T >$ 25 GeV. This energy is the jet energy after all corrections to the jet energy scale have been applied, described in detail in Chap. 5. We order jets in E_T in the event, so the "first" jet is the highest-E_T jet, and the "second" jet the second-highest in E_T. These jets are also required to be within a well-reconstructed portion of the detector, with $|\eta| <$ 2. Furthermore, we require that the jets not overlap with one of the reconstructed leptons by requiring the ΔR between the leptons and any jet to be greater than the cone-size of the jet, 0.4. We also require that the jets be well-separated from each other, with ΔR between jets $>$ 0.7—the increased separation is motivated by the use of a quark-gluon discriminant (see Chap. 7) that looks at tower and track information within a cone of R= 0.7 around the jet.

Finally, because our final state should contain no objects that fail to be reconstructed

in the detector, we also require the missing transverse energy ($\not\!\!E_T$) be less than 20 GeV.

4.3 Background and Signal Modeling

After applying the selection from Tab. 4.1, we have three major background contributions to the sample: events with a $Z \to \ell^+\ell^-$ decay plus additional jets; $t\bar{t}$ production, where the decay chain is $t\bar{t} \to W^+bW^-\bar{b} \to \ell^+\nu_\ell b\ell^-\bar{\nu}_\ell\bar{b}$ (which is suppressed but not eliminated by the $\not\!\!E_T$ cut); and, events where one or both leptons are "faked" by jets that satisfy the electron or muon selection requirements. The former two backgrounds, along with our ZW/ZZ signal, are modeled using Monte Carlo (MC) simulations, while the events with fake leptons are modeled using a data-driven method. The details of the modeling are described below.

4.3.1 Monte Carlo Simulations

The dominant background in our ZW/ZZ search is from the leptonic decay of a Z boson produced in association with two jets. We model this background using MC simulations generated using ALPGEN [49] version 2.10. ALPGEN is a leading-order (LO) matrix element calculator and event generator, focused specifically on modeling multiparton final states. $Z + n$ parton final states are generated separately for varying n, and then these events are interfaced with another MC program, PYTHIA [50] (version 6.216), which performs the evolution of the parton shower into its final state particles. A matching scheme accounts for ambiguities between partons generated in the initial interaction and those produced in the parton shower, avoiding double-counting.

We apply a k-factor of 1.4 to the predicted cross section from ALPGEN to account for next-to-leading-order effects (NLO) that increase the Z + jets cross section, and we apply an additional k-factor of 2.0 to $Z + b\bar{b}$ processes to agree with measurements of the

Z+b jet cross section [51]. While the largest cross section exists for the mediating Z boson to have a mass at m_Z = 91.2 GeV/c^2, we include contributions from Drell-Yan production where the mediator is off-shell. We analyze samples with γ/Z^* mass down to 20 GeV/c^2, and up to 400 GeV/c^2, though the cut on $m_{\ell\ell}$ greatly suppresses any contributions outside our Z mass window.

We use simulations generated using PYTHIA alone to model our $t\bar{t}$ background and our diboson signal samples. Our $t\bar{t}$ sample assumes a top mass of m_t = 172.5 GeV/c^2, very close to the current world average value of 173.5 GeV/c^2 [18]. We assign a production cross section of $\sigma_{t\bar{t}}$ = 7.5 pb. The top sample we use is an inclusive sample containing all possible decays of the W bosons produced, not only the leptonic decays. Our diboson samples are also inclusive decay samples, and include a generated γ/Z^* mass down to 2 GeV/c^2 (the W boson is required to be on-shell). We use the predicted cross sections according to [19], σ_{WZ} = 3.7 pb and σ_{ZZ} = 1.4 pb for $m_{Z^*} > 40$ GeV/c^2 (see [31]), scaling our generated MC accordingly.

All MC samples are generated using the CTEQ6M [52] parton distribution functions. After event generation, the simulated events are run through a GEANT3 [53] simulation of the CDF detector. GFLASH [54] is used to speed-up the simulation of particle showers in the calorimeter. A complete list of the various MC samples used in this analysis is provided in App. A.

Monte Carlo simulations are generated using an instantaneous luminosity profile of a particular run range: there are “low-luminosity” samples that correspond to early periods of data-taking, while “high-luminosity” samples were generated mimicking later data, when the instantaneous luminosity provided by the Tevatron was higher. Still, the luminosity profile of our MC samples does not perfectly match that of our data. We account for this by applying a weight to MC events according to the number of reconstructed interaction vertices, N_{vtx}, a stand-in for the instantaneous luminosity (and a variable that captures

the effects of pileup of multiple interactions). This reweighting factor is determined by comparing the distribution of N_{vtx} in data and MC in a sample of $Z \to \ell^+\ell^-$ events before any cuts on the jets in the event or the $\not\!\!E_T$ have been applied, so as not to bias the N_{vtx} distribution.

4.3.2 "Fake" Leptons

Another significant background results from jets that pass the lepton selection cuts described in Secs. 3.1 and 3.2. These jets are thus misidentified as leptons, and we call them "fake" leptons, or simply "Fakes". The contributions from these lepton fakes are estimated via a data-driven method, but this method is different for muons and electrons. For the former, we use events with same-sign muon pairs (rather than opposite-sign) that otherwise satisfy all of our event selection requirements. The reconstructed charge for jets faking muons should be uncorrelated with the charge of another muon in the event, so the contribution from same-sign fakes should be the same as that of opposite-sign fakes. Since we expect no major physics backgrounds to our same-sign sample, we may simply use it to model our muon fakes.

In order to estimate the number of events in our sample where a jet fakes an electron, we construct a "fake rate" that corresponds to the likelihood a jet with a certain E_T and η will fake our electron selections. To determine this fake rate, we look at data from three jet triggers: `Jet50`, `Jet70`, and `Jet100`, where the trigger object is a calorimeter jet with E_T greater than 50 GeV, 70 GeV, and 100 GeV, respectively. The former two triggers have a pre-scale applied.

In these samples, we search for all jets with $|\eta| < 2.8$ (which matches the coverage of PEM objects) and EM fraction < 0.9 in bins of E_T and η. The number of such jets forms the denominator of our fake rate. For the numerator, we count all identified leptons matched

to a jet (where we drop the EM fraction cut). In all cases, we ignore the highest energy jet in the event, so as not to be biased by the trigger. Also, we reduce the number of real electrons in these events from $W^{\pm} \to e^{\pm}\nu$ and $Z \to e^{+}e^{-}$ decays by removing events with $\not\!\!E_T > 15$ GeV (removing W decays) and those with more than one identified lepton (removing Z decays). Remaining contributions from W and Z production are modeled with PYTHIA Monte Carlo samples, and are subtracted from the numerator of our fake rate. The overall fake rate is then

$$\mathcal{F}(E_T, \eta) = \frac{(N_e^{\text{Data}} - N_e^{MC})(E_T, \eta)}{N_{\text{jets}}(E_T, \eta)} \ , \tag{4.1}$$

where N_e^{Data} is the number of "electrons" matched to jets in the jet-triggered data (and thus are largely fake electrons), N_e^{MC} is the number of electrons expected from W and Z production, and N_{jets} is the total number of jets in the jet-triggered data.

The fake rates calculated with data from the three jet triggers, integrated over η, are shown in Figure 4.1. In our analysis, we use the `Jet70` fake rates, assigning a 50% systematic uncertainty to cover any discrepancies between triggers. These differences may be from some residual trigger bias, or from some difference in the quark and gluon content of the samples. We expect quark jets to be more likely to fake electrons because they are more collimated than gluon jets.

To use our fake rates to estimate the number of fake electrons in our sample, we identify e-jet pairs (instead of e-e pairs; the jets are required to have $|\eta| < 2.8$ and EM fraction < 0.9) in the high-p_T electron dataset, and assign $\mathcal{F}(E_T^{\text{jet}}, \eta^{\text{jet}})$ as a weight to these e-jet pairs. After this weight is applied, we treat that jet as we would any other electron, and then require the event to pass all requirements in Tab. 4.1. We consider each e-jet pair multiple times, applying the the fake rate for all electron types that may apply to the jet. Also, since there may be multiple jets in an event, we consider all possible e-jet pairs

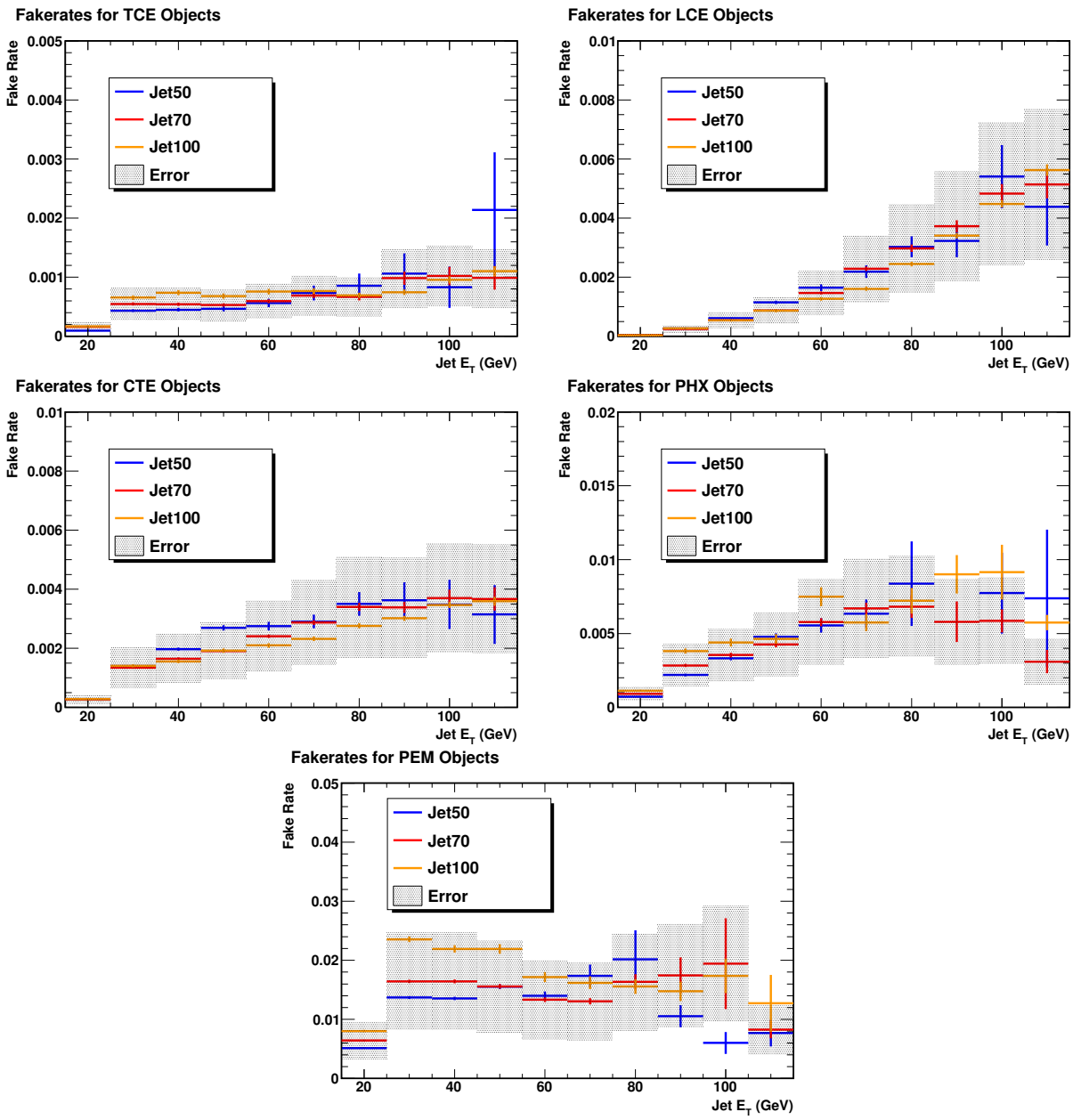

Figure 4.1: Fake rates calculated from jet data for our electron categories (from top left, moving right and down: TCE, LCE, CTE, PHX, and PEM). The shaded region corresponds to a 50% uncertainty on the `Jet70` fake rates.

in the event.

The jets are treated exactly like electrons with one exception: we scale the energy of the jet down, because if a jet with cone-size $R = 0.4$ and a given energy fakes an electron,

that electron will have a smaller energy than the original jet. We determine this energy scaling by comparing jet energies and the matched electron energies in the jet data, fitting the electron E_T as a linear function of the jet E_T with a turn-on curve near 20 GeV:

$$E_T^{\text{electron}} = \frac{(a \cdot E_T^{\text{jet}} - b)}{(1 + e^{-c \cdot (E_T^{\text{jet}} - d)})} + 20 \text{ [GeV]} . \tag{4.2}$$

These relationships are shown in Fig. 4.2. Some jets with low energies look like they are not modeled well in the fit, but this is largely due to low statistics as it is unlikely for a 20 GeV jet to fake a 20 GeV/c lepton. Such jets are assigned fake rates at practically zero, therefore their contribution is negligible.

The method described above for applying the fake rates to e-jet pairs works well for modeling events where there is one real electron (likely from a leptonic W decay) and one fake electron. However, it does not correctly model the number of events where both electrons are faked by jets. To illustrate this, suppose we have an event with N_j jets. Ignoring the complications of the E_T/η binning, the likelihood exactly one electron being faked by this set of jets is

$$N_1 = \mathcal{F} \cdot N_j . \tag{4.3}$$

The likelihood of two electrons being faked has a combinatoric factor applied:

$$N_2 = \mathcal{F}^2 \cdot \frac{N_j!}{2 \cdot (N_j - 2)!} = \mathcal{F}^2 \cdot \frac{N_j(N_j - 1)}{2} . \tag{4.4}$$

Comparing Eq. 4.3 and Eq. 4.4, it's easy to see that $N_2 \neq \mathcal{F} N_1$, which is what the modeling described previously assumes. Thus, for events where the two electrons are both fake electrons from jets, we must apply an additional factor of $(N_j - 1)/2$. Note that here, N_j is the total number of jets in an event, assuming all electrons are also jets. Thus, our selection in Tab. 4.1 requires $N_j \geq 4$. In truth, our selection has $< N_j >= 4.3$, thus would

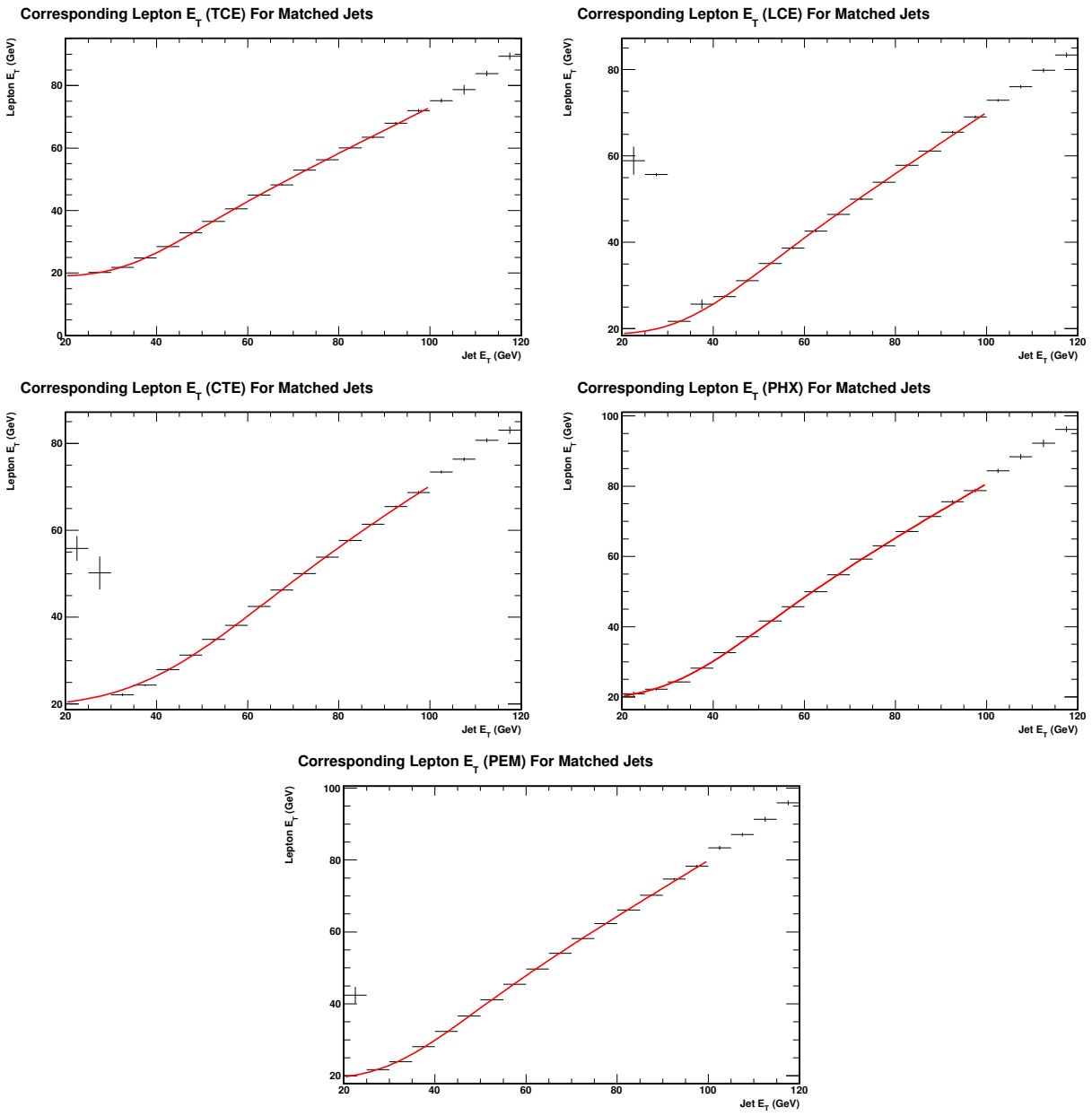

Figure 4.2: The matched electron E_T as a function of the jet E_T in `Jet70` data, used to determine the approximate scale of jet energy to their faked lepton energy in the high-p_T electron datasets. From top left moving right and down: TCE, LCE, CTE, PHX, and PEM. The red line represents a fit to the data. Some bins at low jet energies look like they are not modeled well in the fit, but these bins contain a small number of events, and thus are assigned fake rates at practically zero, therefore their contribution is negligible.

need a correction factor of ≈ 1.65, assuming all fake electron events come from these double-fake events. This turns out to be a good assumption when requiring $\not\!\!E_T < 20$ GeV, based on comparisons of data and W+ jets MC.

4.4 Scale Factors from $Z + 1$ Comparisons

Our Monte Carlo simulation does not always accurately model important factors that affect the number of data events we see. For instance, the MC does not have a trigger applied to it, and so a non-unity trigger efficiency can lead to fewer data events than is modeled by the MC. Also, the MC may not correctly model the identification and reconstruction efficiencies for an electron or muon to fall into one of the definitions in Secs. 3.1 and 3.2. While it is possible to measure these quantities independently, our holistic trigger selection and the variety of lepton definitions make it simpler to compare data and MC in a separate sample dominated by background events.

$Z + 1$ jet Selection, for Lepton-Pair Scale Factors
$N_{vert} > 0$
$\lvert z_0 \rvert < 60$ cm
Δz_0 between leptons < 5 cm*
Leptons oppositely charged*
$1^{st}/2^{nd}$ lepton $p_T > 20$ GeV/c
76 GeV/$c^2 < M_{\ell\ell} < 106$ GeV/c^2
Z $p_T > 10$ GeV/c
$\not\!\!E_T < 20$ GeV
$N_{\text{jets}} > 20$ GeV = 1
$\lvert \eta_{\text{jet}} \rvert < 2.4$
ΔR between jet and any lepton > 0.4

Table 4.2: Summary of event selection for determining lepton-pair scale factors.

*Only applied to central electron and muon pairs.

We determine these scale factors in a $Z + 1$ jet selection. The requirements for this selection are similar to those in Tab. 4.1, and are summarized in Tab. 4.2. The resulting

scale factor to have the MC agree with the data is then

$$k_{\mathrm{SF}} = \frac{N_{\mathrm{data}} - N_{\mathrm{fakes}}}{N_{\mathrm{MC}}} ,$$

where N_{data}, N_{fakes}, and N_{MC} are the number of events in data, the estimated number of fake lepton events, and the estimated number of events from MC, respectively. The scale factors are determined per lepton-pair, and listed in Tab. 4.3. These scale factors are applied as a weight to each MC event individually.

4.5 $Z + 2$ Jet Signal Region Comparisons

We apply the selection requirements in Tab. 4.1 to both the data and background selection. The total number of events after selection is summarized in Tab. 4.4. Overall, we see good agreement between the data and MC, with a difference in the total number of expected and observed events at less than 2%.

4.5.1 Comparisons of Leptonic Variables

Figs. 4.3-4.6 show distributions of some of the leptonic variables. We include absolute comparisons alongside comparisons where the number of events has been normalized to unity, in order to better compare the shape. In the final fit for signal events, the Z + jets cross section will be allowed to vary unconstrained, and so may move to cover any discrepancy in the total number of expected events. Thus, these shape comparisons are in many ways more indicative of the agreement between data and our modeling in terms of the final result. We see some mismodeling of the resolution of lepton energies, evident in the comparison of the Z mass, due to small muon track p_T resolution differences between data and MC. Overall the data and MC agree well.

Electron-Pair Categories	k_{SF}
TCE-TCE	1.090
TCE-LCE	1.108
TCE-CTE	1.091
TCE-PHX	1.012
TCE-PEM	1.153
TCE-CrkTrk	0.847
LCE-CrkTrk	0.774
CTE-CrkTrk	0.733

Muon-Pair Categories	k_{SF}
CMUP-CMUP	0.813
CMUP-CMU	1.150
CMUP-CMP	0.810
CMUP-CMX	0.915
CMUP-CMIOCES	0.992
CMUP-CMIOPES	0.716
CMUP-CrkTrk	0.830
CMU-CMU	0.725
CMU-CMP	0.756
CMU-CMX	1.205
CMU-CMIOCES	0.678
CMU-CMIOPES	0.405
CMU-CrkTrk	0.464
CMP-CMP	0.552
CMP-CMX	0.875
CMP-CMIOCES	0.537
CMP-CMIOPES	0.332
CMP-CrkTrk	0.414
CMX-CMX	0.985
CMX-CMIOCES	1.119
CMX-CMIOPES	0.760
CMX-CrkTrk	0.837
CMIOCES-CMIOCES	0.280
CMIOCES-CMIOPES	0.123
CMIOCES-CrkTrk	0.165

Table 4.3: List of lepton category pairs used in this analysis, along with the necessary Monte Carlo scale factors that incorporate trigger efficiencies and ID reconstruction efficiencies. Categories which do not require oppositely charge leptons are italicized. See Sec. 3.1 and 3.2, which contains a description of each lepton definition.

4.5.2 Comparisons of Jet Variables

Distributions of some jet variables are shown in Figs. 4.7-4.12. We generally see good agreement between data and MC, taking into account the slight normalization differences between the two. There are a couple items of note.

- The dijet mass distribution (see Fig. 4.7), which we fit to in order to extract our ZW/ZZ signal, is reasonably well-modeled, but the data has a somewhat broader

	e-e Events	μ-μ Events	All Events
Z + jets	5216 ± 670	3609 ± 463	8825 ± 1133
Z + b jets	417 ± 175	311 ± 130	728 ± 305
$t\bar{t}$	5.76 ± 0.56	3.98 ± 0.39	9.74 ± 0.95
Fakes	348 ± 174	3 ± 1.7	351 ± 174
Expected ZW/ZZ	184 ± 17	131 ± 12	315 ± 29
Total Expected	6171 ± 774	4058 ± 509	10228 ± 1284
Data	5990	4049	10039

Table 4.4: The number of events in the Z + 2 jet signal region. The uncertainties are described in further detail in Sec. 8.2. In this table, we take an uncertainty on the Z + jets and ZW/ZZ cross section of 10% and 6%, respectively.

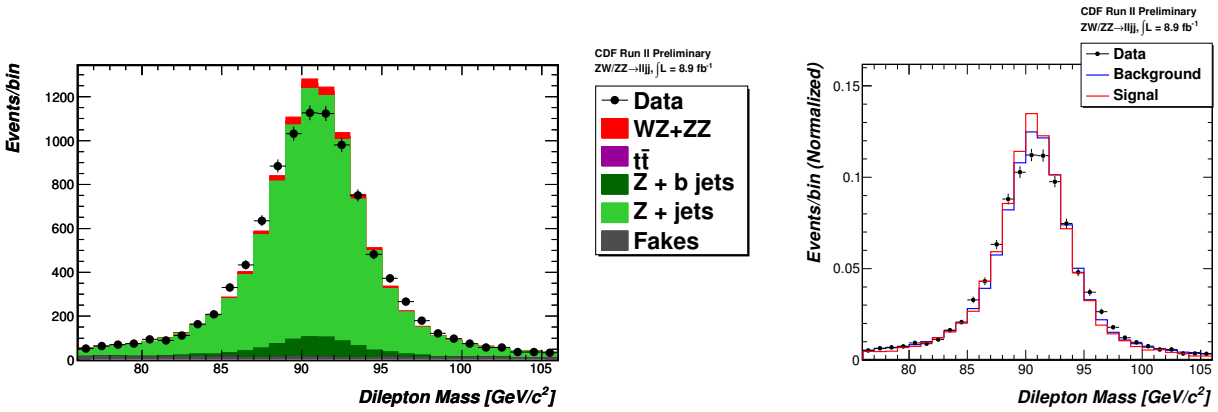

Figure 4.3: The distribution of dilepton mass in the Z + 2 jet signal region, in linear scale (left) and normalized (right). The number of events from each contributing sample in the prediction, and the number of observed data, is given in Table 4.4.

dijet mass spectrum than the background model.

- The model of the E_T of the jets (Fig. 4.8) agrees well at high jet E_T, but has a slight deficit at lower jet energies.

- The most significant mismodeling we see is in the η of the sub-leading jet (Fig. 4.9), where we see a larger number of forward jets in data than in our background model. However, this discrepancy is not present in the leading jet, and is rather minor. Other kinematic variables related to the jet η, like the ΔR between jets (Fig. 4.10), are well-

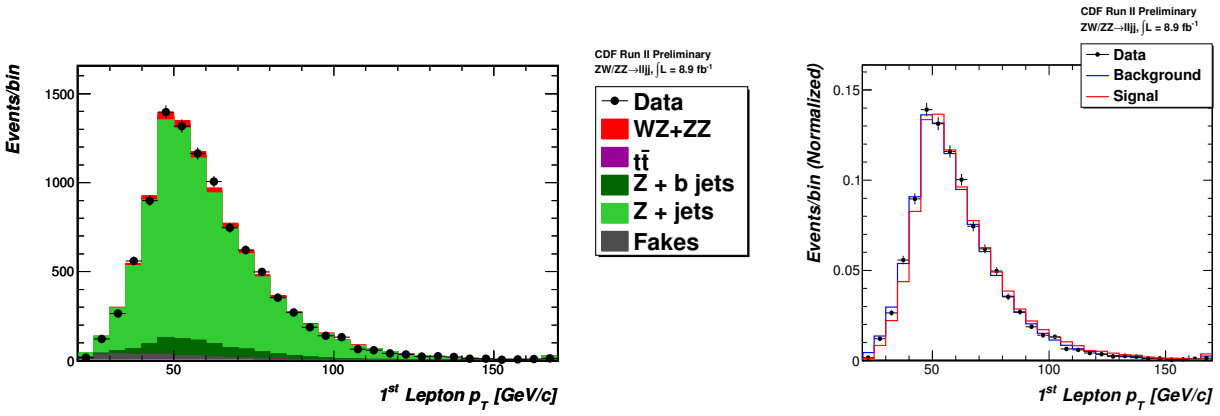

Figure 4.4: The leading lepton p_T in the $Z + 2$ jet signal region, in linear scale (left) and normalized (right).

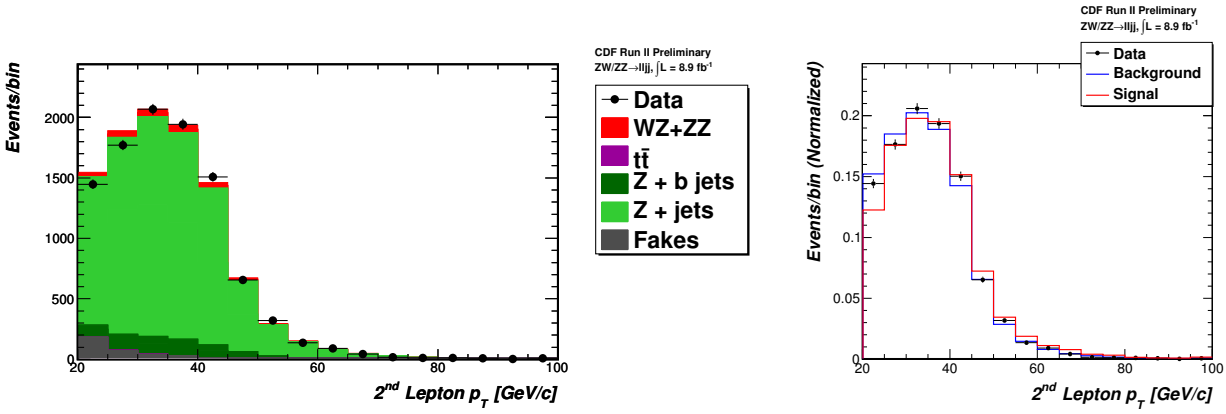

Figure 4.5: The sub-leading lepton p_T in the $Z + 2$ jet signal region, in linear scale (left) and normalized (right).

modeled.

- The dijet p_T (Fig. 4.12), is particularly well-modeled.

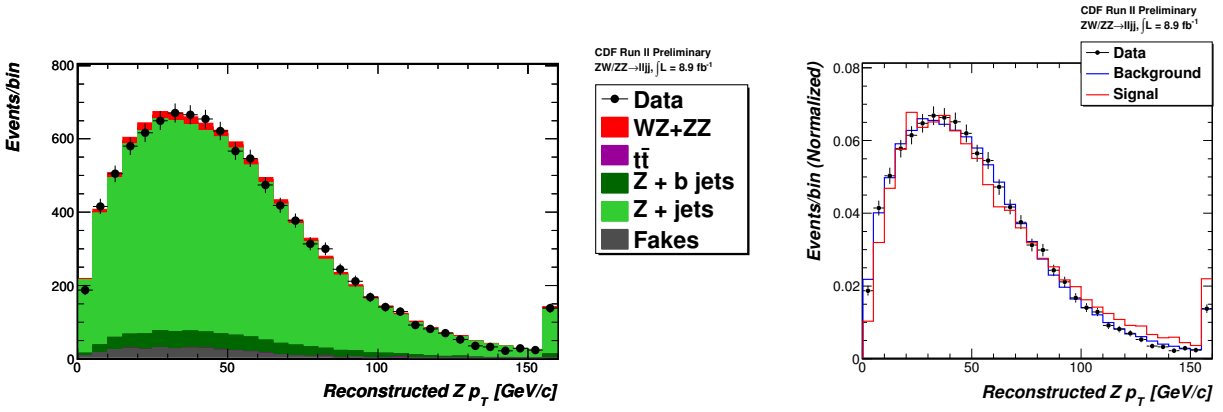

Figure 4.6: The reconstructed Z p_T in the $Z + 2$ jet signal region, in linear scale (left) and normalized (right).

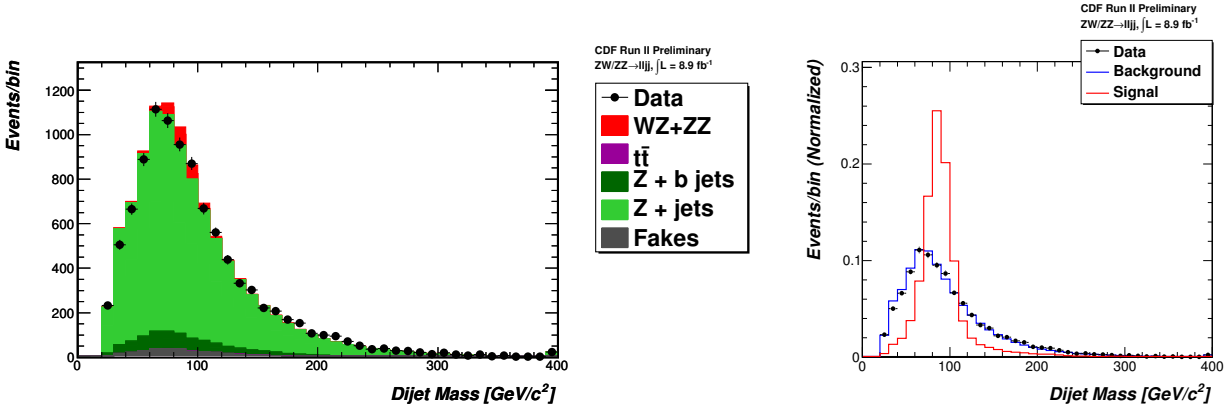

Figure 4.7: The distribution of dijet mass in the Z + 2 jet signal region, with MC scaled to the data's luminosity (left) and with the distributions normalized to unity (right).

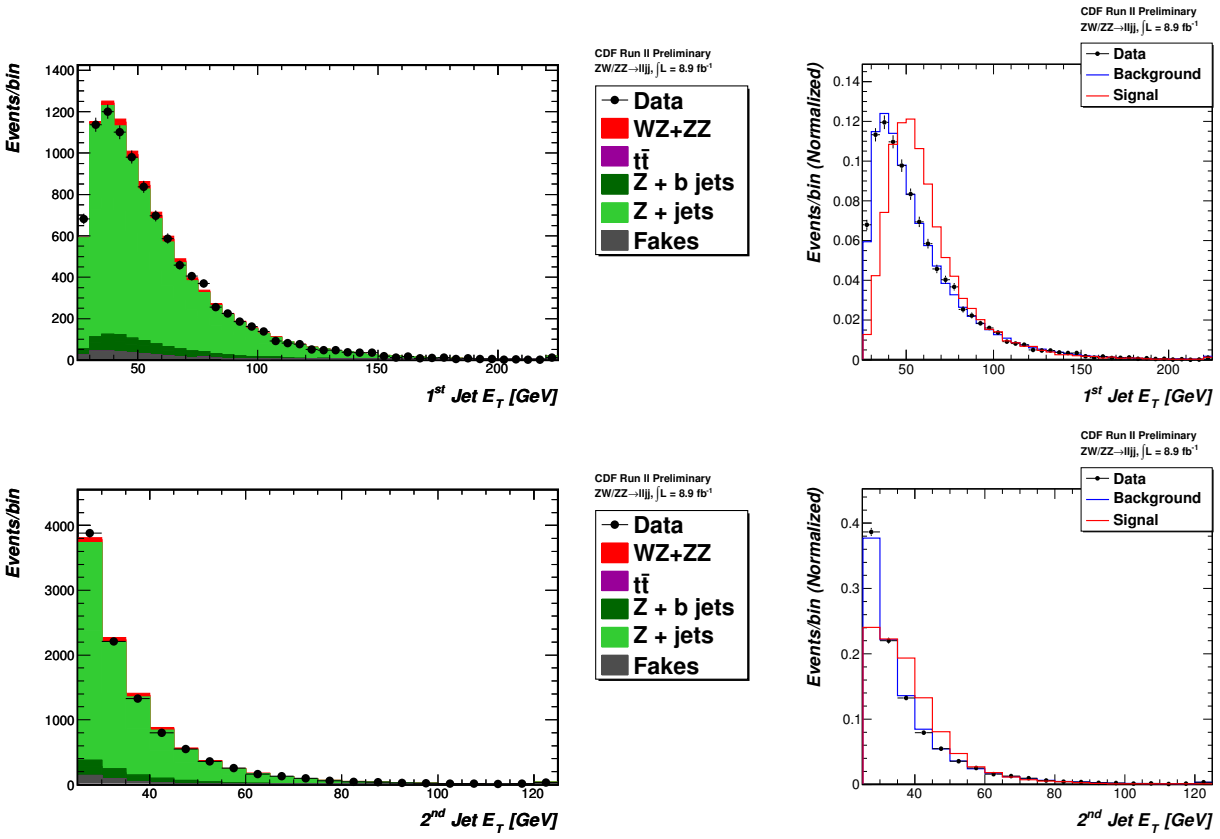

Figure 4.8: The distribution of the leading jet E_T in the Z + 2 jet signal region, with MC scaled to the data's luminosity (top left) and with the distributions normalized to unity (top right), and the second leading jet E_T (bottom row).

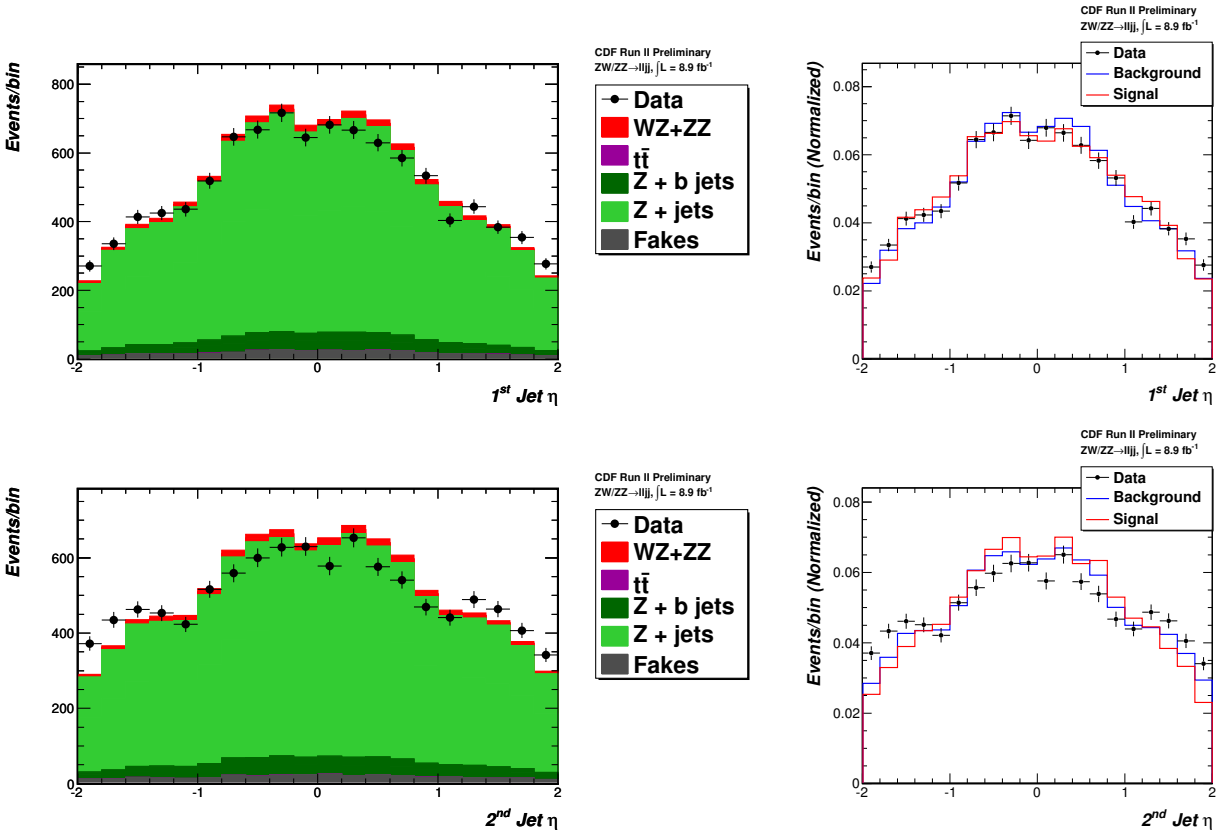

Figure 4.9: The distribution of the leading jet η in the Z + 2 jet signal region, with MC scaled to the data's luminosity (top left) and with the distributions normalized to unity (top right), and the second leading jet η (bottom row).

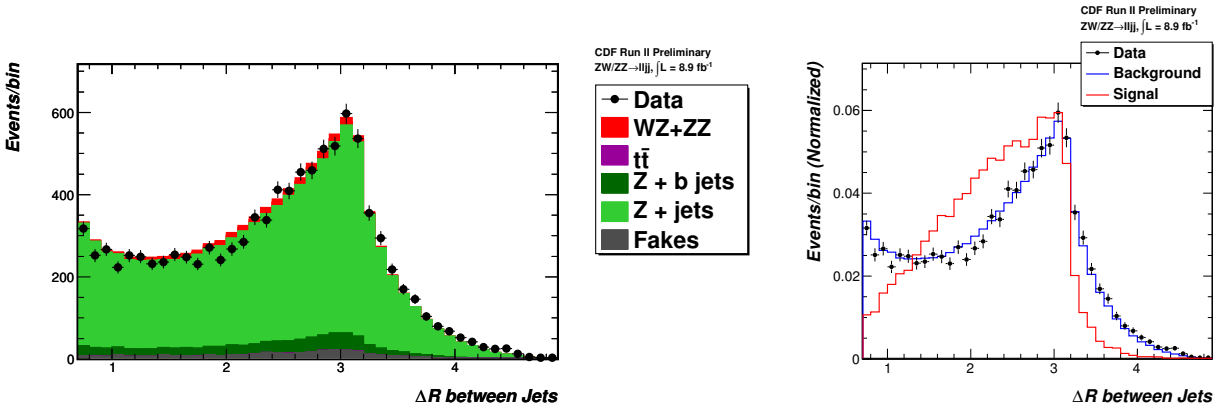

Figure 4.10: The distribution of ΔR between jets in the Z + 2 jet signal region, with MC scaled to the data's luminosity (left) and with the distributions normalized to unity (right).

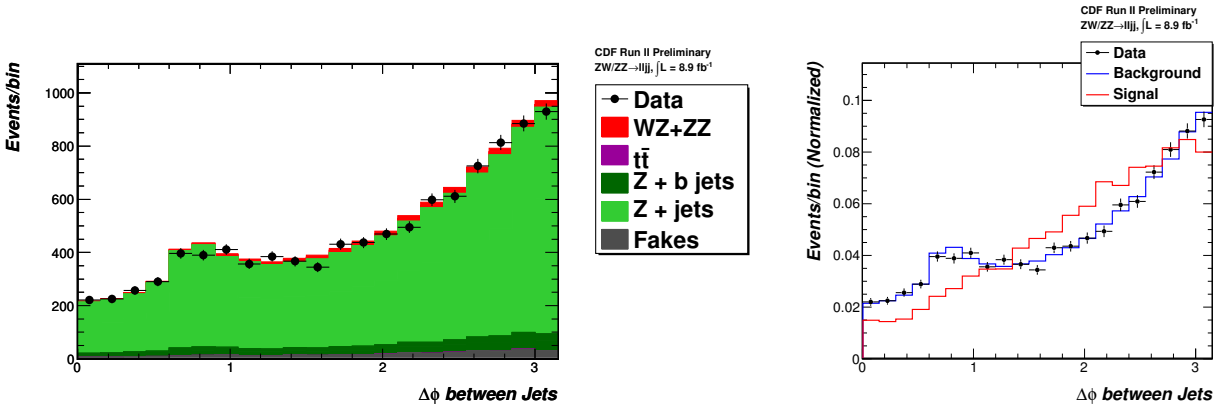

Figure 4.11: The distribution of $\Delta\phi$ between jets in the Z + 2 jet signal region, with MC scaled to the data's luminosity (left) and with the distributions normalized to unity (right).

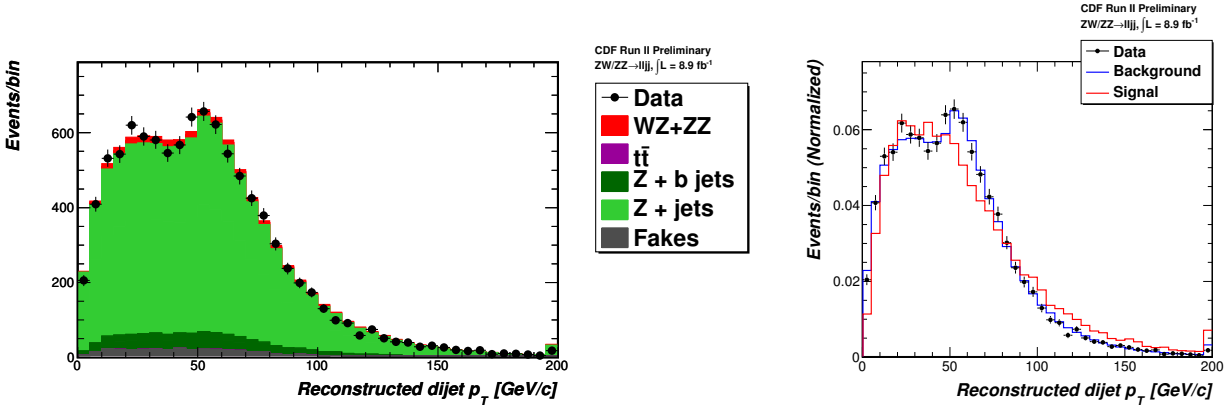

Figure 4.12: The distribution of dijet p_T in the Z + 2 jet signal region, with MC scaled to the data's luminosity (left) and with the distributions normalized to unity (right).

CHAPTER 5

JET ENERGY CORRECTIONS

Hadronic jets resulting from the fragmentation of quarks and gluons are observed as clusters of energy in the calorimeters. The jet clustering algorithm is described in further detail in Sec. 3.3. However, many effects alter the measured jet energy so that it is different from the actual energies of the particles making up the jet, and different from the original energy of the parton that created the jet. The goal of the jet energy corrections at CDF is to take the initial jet energies and scale them so that they match the energy of the particles that make up the jet, or so that they match the energy of that originating parton. A cartoon outlining this process is shown in Fig. 5.1. In short, the parton's transverse momentum may be written as a series of corrections to the original jet p_T:

$$p_T{}^{\text{parton}} = \left\{ \left[p_T{}^{\text{jet}} \times C_\eta(p_T{}^{\text{jet}}, \eta) - C_{\text{MI}}(N_{\text{vtx}}) \right] \times C_{\text{Abs}}(p_T{}^{\text{jet}}) - C_{\text{UE}} \right\} \times C_{\text{OOC}}(p_T{}^{\text{jet}})$$

where

- C_η is an η-dependent correction that normalizes the calorimeter response over its entire η coverage;
- C_{MI} is a correction for the existence of pileup in an event due to multiple $p\overline{p}$ interactions per bunch crossing, based on the number of reconstructed interaction vertices;
- C_{Abs} is an "absolute" energy scale correction that takes the calorimeter jet energy and scales it to the corresponding energy of the particles that make up the jet;

- C_{UE} is a correction due to the energy deposited in the calorimeter from the soft initial-state radiation or from spectator quarks of the primary $p\overline{p}$ collision; and,

- C_{OOC} is an "out-of-cone" correction that accounts for energy associated with the originating parton that lies outside the fixed cone-size of our jets.

We describe these energy corrections in the following chapter (further detail on the calorimeter calibration and the above corrections is in [48]). In addition, we also derive a new correction for jets simulated in Monte Carlo based on whether the jet comes from a quark or gluon. We describe the necessity for and determination of this jet correction in Sec. 5.6.

5.1 Calorimeter Energy Scale Calibration and Stability

Before any jet energy corrections are applied, we ensure that each of the calorimeters is calibrated to maintain a stable response over time. The CEM energy scale is set so that the Z boson mass measured from well-reconstructed $Z \rightarrow e^{+}e^{-}$ decays agrees with the mass measurement from LEP [56]. The PEM scale is set using using the reconstructed mass from $Z \rightarrow e^{+}e^{-}$ decays with one electron in the CEM and the other in the PEM. The energy scale of the hadronic calorimeters (CHA, WHA, and PHA) is set by their response to 50 GeV/c charged pions based on test beam data, using pions which do not interact in the electromagnetic calorimeter sections. The total energy scale for each calorimeter tower is simply the addition of the properly scaled electromagnetic and hadronic tower energies.

The stability of the calorimeter is checked both online and offline. Generally the energy scale of the calorimeters decreases over time due to aging of the scintillators and phototubes, but after calibrations the energy scale is kept stable. The CEM is monitored over time by the checking the E/p for central electrons; the CHA and WHA are monitored using a laser system, muons fom $J/\psi \rightarrow \mu^{+}\mu^{-}$ decays, and from data taken with

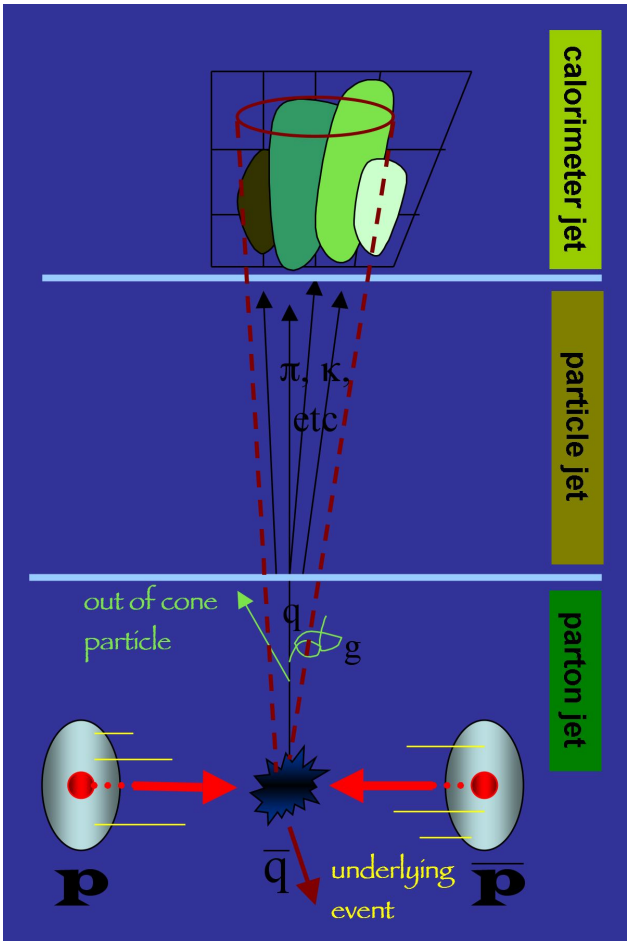

Figure 5.1: A cartoon [55] showing the various levels at which a jet's energy may be represented. We measure clusters of energy in the calorimeter, and so after calibrations and detector-dependent corrections, we can obtain properly measured calorimeter jets. We may associate these calorimeter jets with the particles that make them up, which together form a particle jet. Finally, when taking into account effects that may leave energy outside the cone of the jet, we can correct the energy to the parton level.

a minimum bias trigger; and, the PEM and PHA calorimeters are monitored with a laser system and a radioactive Co^{60} source. Independent $Z \rightarrow e^+e^-$ and $W \rightarrow \mu\nu_\mu$ samples verify the electromagnetic and hadronic calorimeters' stability, respectively (additionally, the jet response in the PHA is used to check stability there). The uncertainty on CEM and CHA stability is 0.3% and 1.5%. Thus, for a central jet which leaves 70% of its energy in the CEM and 30% in the CHA, the uncertainty on the jet energy due to the calorimeter stability is 0.5% [48].

5.2 η-dependent Corrections

After the calibrations on the calorimeters' energy scales, the response of the CDF calorimeter is still not uniform in η. There are two main reasons for this. First, there are significant gaps in the coverage of the calorimeter across η. The central calorimeters were constructed in two halves, and where the halves come together at η = 0 there is a dead region. Similarly, there is a gap between the central and plug calorimeters at $\eta \approx \pm 1.1$ which leads to a loss in coverage. Thus, jets located within a ΔR = 0.4 of these regions will have a lower measured energy than jets in other parts of the calorimeter. The second reason for η dependencies comes from a difference in response between the plug and central calorimeters. The η-dependent corrections account for these effects, and normalize the calorimeter response to match that in the region $0.2 < |\eta| < 0.6$.

The corrections are obtained using a "dijet balancing method". The idea behind all balancing methods is to identify events that contain only two objects: one well-measured, and one "probe" object. If there is little other activity in the event, then these two objects should balance in the transverse plane: *i.e.* the $\Delta\phi$ between them should be nearly π, and that the magnitude of their p_T's should be equal. Thus, we know that the p_T of the probe object should be the same as the well-measured one, and we can determine corrections

on the energy of the probe object to make that the case.

For the determination of the η-dependent corrections, the well-measured object is a "trigger jet", a jet with $0.2 < |\eta| < 0.6$. This jet is required to back-to-back with a second, probe jet: $\Delta\phi > 2.7$ rad. If a third jet is present in the event, it is required to have very little energy (10 GeV or less, depending on the sample). There must also be no significant $\not\!\!E_T$ in the event. Different corrections for data and MC are derived: in data, the corrections come by looking at dijet balancing in a series of jet-triggered data, where the corrections for higher E_T jets come from higher-E_T jet triggered data; for MC, the corrections come from a dijet sample generated using PYTHIA.

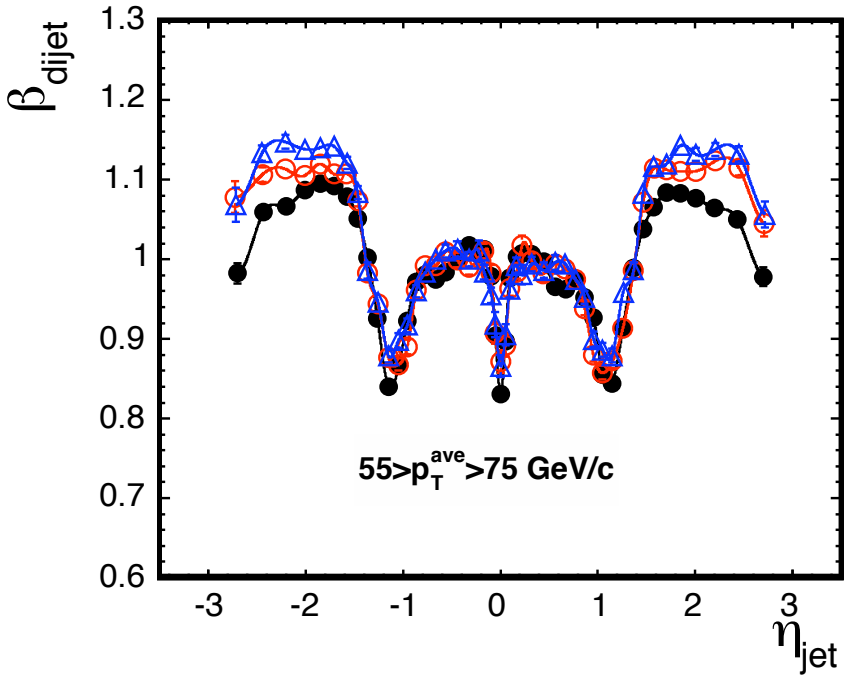

Figure 5.2: The dijet balance $\beta_{dijet} = {p_T}^{\text{probe jet}}/{p_T}^{\text{trigger jet}}$ in the dijet sample used to determine the η-dependent jet energy corrections. Here we show jets in data (black), PYTHIA MC (red), and HERWIG MC (blue) [57] where the average p_T of the probe and trigger jets is between 55 GeV/c and 75 GeV/c. Other p_T ranges are shown in [48].

Fig. 5.2 shows an example of the dijet balance $\beta_{dijet} = {p_T}^{\text{probe jet}}/{p_T}^{\text{trigger jet}}$ as a function

of the probe jet's η in both data and MC for events where the average p_T of the probe and trigger jets is between 55 GeV/c and 75 GeV/c. Corrections for jets are parameterized in η and jet p_T. The uncertainties on this correction come from limitations in the η and p_T parameterization, and from observed differences when varying the cuts on the energy of the third jet and the $\not\!\!E_T$-significance. The uncertainties on the η-dependent corrections of jets with $|\eta| < 2.0$ are typically $\sim$ 1.5% for jets with $p_T \sim 20$ GeV/c, and $\sim$ 0.5% for jets with $p_T > 55$ GeV/c.

5.3 Multiple $p\overline{p}$ Interaction Correction

Additional $p\overline{p}$ collisions may happen alongside a primary interaction in each bunch crossing. Hadrons produced in these additional $p\overline{p}$ collisions, often called "pileup", may end up in jets from the primary interaction, and thus increase the measured energy of the jets. The amount of pileup increases with the instantaneous luminosity, which is constantly changing during a store. In order to have consistent energy measurements independent of the instantaneous luminosity, we must correct for the effect of pileup.

We parameterize the amount of pileup using the number of reconstructed z-vertices, N_{vtx}. We then measure the average transverse energy in a cone from a minimum bias data sample, and determine this average for different N_{vtx}. The results in cones of size $R = 0.4$ are shown in Fig. 5.3. The cone is centered around a randomly selected seed tower in the η range $0.2 < |\eta| < 0.6$. The amount of energy per interaction vertex is roughly linear, and is parameterized with a straight line: $\langle E_T \rangle = 0.006 + 0.356 \times N_{\text{vtx}}$ [GeV]. (The intercept is not exactly zero, as we would expect it to be, due to inefficiencies in finding vertices.) Therefore, each additional reconstructed vertex adds about 356 MeV to each jet in the event, an energy we can then subtract out. The uncertainty on this correction, coming from cross-checks in other data samples, is 15% $\approx$ 50 MeV.

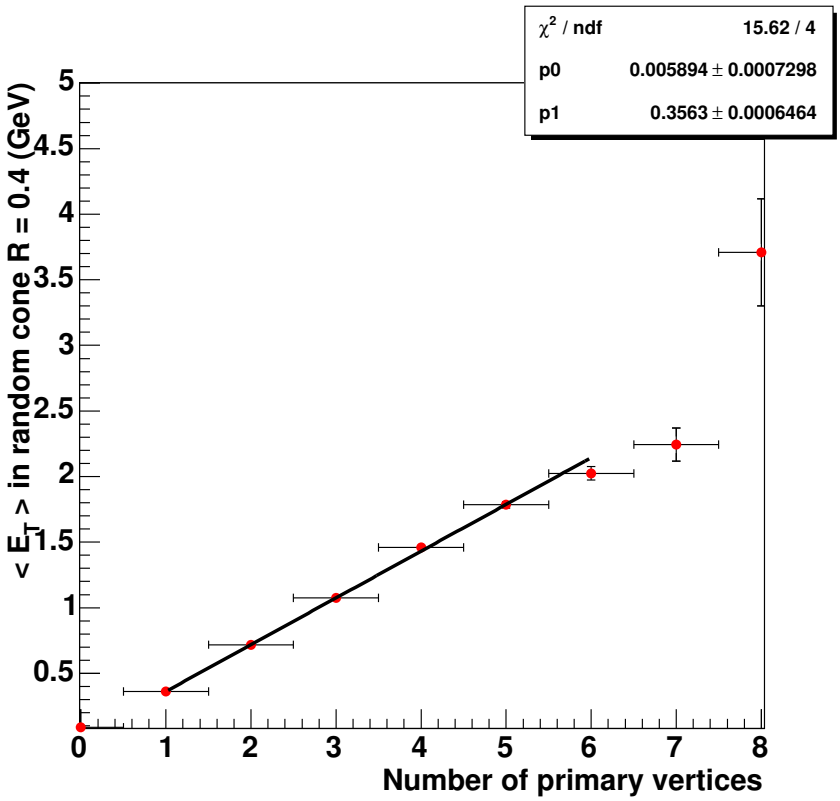

Figure 5.3: The average E_T in a jet with cone size R = 0.4 due to multiple interactions.

5.4 Absolute Jet Energy Corrections

With accurate measurements of the energies of calorimeter jets, we then would like to correct the jet energies so that they better reflect the energies of the stable particles produced in the hadronization. These particle jets are closer to the actual physical quantities we would like to measure, and should be independent of properties of the CDF detector. Thus, the energy scale for these particle jets is called the "absolute jet energy scale".

Because we cannot reconstruct or accurately determine all of the particles in a jet in the CDF detector, we rely on Monte Carlo simulations to obtain this correction. Using a PYTHIA dijet MC sample, we obtain a probability density function for the probability of observing a calorimeter jet with ${p_T}^{\text{jet}}$ given a particle jet of ${p_T}^{\text{particle}}$. Stable particles (those with lifetimes long enough to hit the calorimeters) are clustered into jets in a similar way to how calorimeter towers are clustered, described in Sec. 3.3. The probability density function is then parameterized as a double-Guassian that is a function of $\left({p_T}^{\text{particle}} - {p_T}^{\text{jet}}\right)$. An unbinned likelihood fit is used to determine the free parameters of the double-Gaussian, which are allowed to vary linearly in ${p_T}^{\text{particle}}$. The jets used to determine this correction lie in the region $0.2 < |\eta| < 0.6$, and are required to have the location of the centers of their particle and calorimeter jets be within $\Delta R = 0.1$.

The absolute energy corrections for cone-size $R = 0.4$ jets are shown in Fig. 5.4. For low-p_T jets, the correction is rather high – 20 GeV calorimeter jets are typically undermeasured with respect to the actual energy of the particles that produce them by about 35%. The correction is lower for higher-p_T jets, approaching a constant value of about 1.12.

While the uncertainties for this correction in MC, where we can directly link the calorimeter and particle jets, are negligible, an uncertainty in the correction arises from the fact that the particle jets in MC may not accurately represent those in data. We include uncertain-

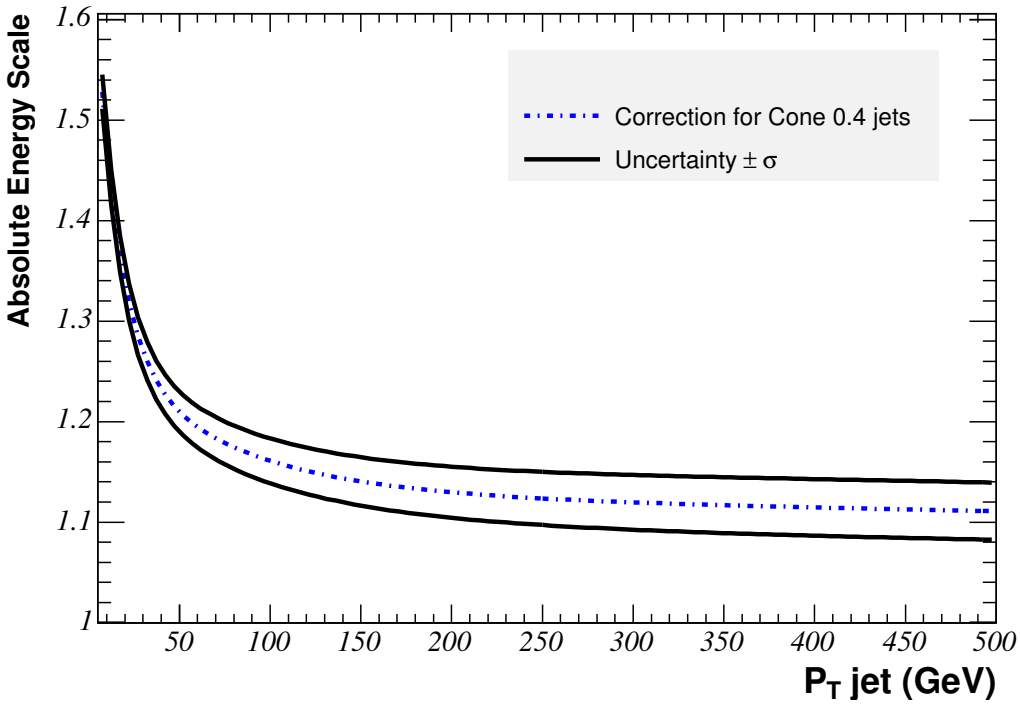

Figure 5.4: The absolute energy correction as a function of p_T^{jet}. The uncertainties are shown in more detail in Fig. 5.5.

ties from possible differences in the single-particle response in both the electromagnetic and hadronic calorimeters, taken from measurements of that response in minimum bias events with single isolated tracks pointing into the calorimeters. We also include uncertainties in the fragmentation model, which may lead to differences in both the particle multiplicity and momentum spectrum. An additional uncertainty arises from the calorimeter stability, which is known to 0.5%. The individual and total systematic uncertainties for the absolute jet energy scale are shown in Fig. 5.5.

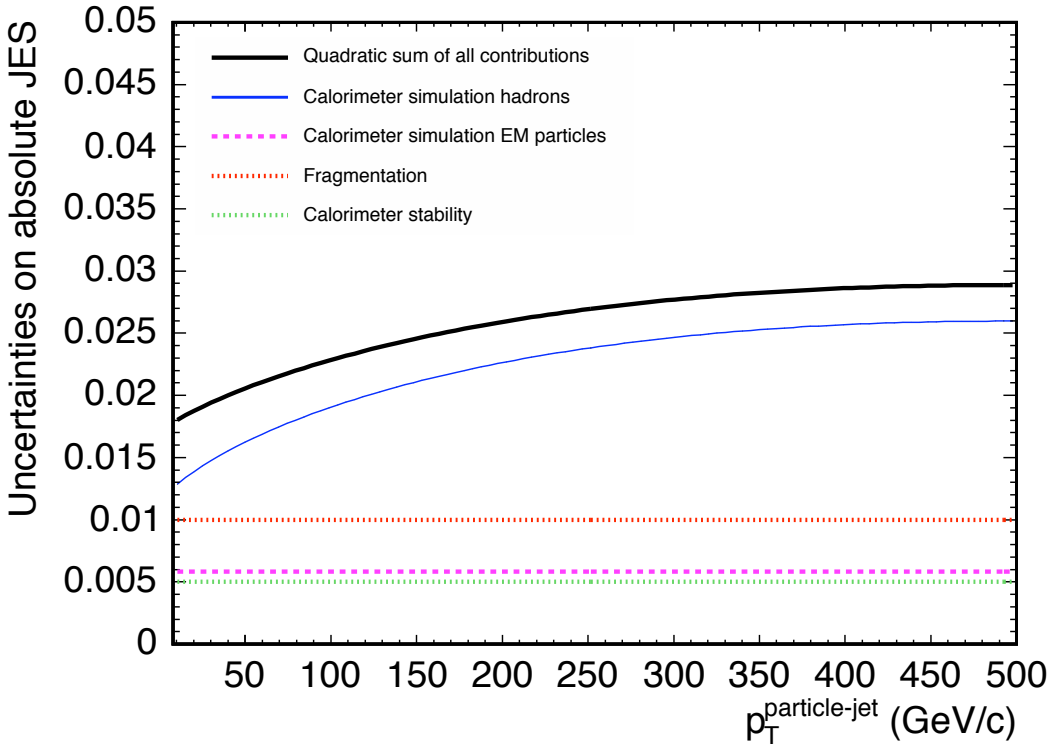

Figure 5.5: The systematic uncertainties on the absolute energy correction as a function of p_T^{jet}.

One thing important to point out here: the derived corrections to the absolute energy scale do not vary based on the original parton type. Jets originating from quarks and those originating from gluons have distinct differences in their width, particle multiplicity, particle momentum spectrum, and other variables that may affect the energy scale [58].

This serves as a motivation for corrections to quark and gluon jets separately, but it is impossible to do this in the data where, *a priori* (and often *a posteriori*), a jet's progenitor cannot be determined. Instead, we derive separate corrections to MC jets only based on their originating parton. These corrections come after all other corrections, and are described in Sec. 5.6.

5.5 Underlying Event and Out-of-Cone Energy Corrections

While the corrections to the absolute energy scale allow for comparisons of jets independent of the CDF detector, the most fundamental quantity to reconstruct from a jet is the original energy of the parton that created it. With this energy accurately determined, we should obtain the best measurements for quantities like the invariant mass of a dijet system. Corrections from the absolute energy scale to the parton energy need to account for two main effects: first, additional energy reconstructed in jets coming from other elements of the $p\overline{p}$ collision, like soft radiation from the collision's hard-scattering partons, or particles from interactions of spectator partons; second, energy lost outside the cone of the reconstructed jet, due to final state radiation from a parton or from hadronization that extends outside the cone size of the jet. Correcting for this underlying event and out-of-cone element should bring us closer to the initial energy of the parton.

We derive these corrections simultaneously in a manner similar to the determination of the absolute energy corrections. Using a PYTHIA dijet MC sample, we obtain a probability density function for the probability of observing a particle jet with ${p_T}^{\text{particle}}$ given a parton with ${p_T}^{\text{parton}}$. This probability density is parameterized as a double-Gaussian function of $\left({p_T}^{\text{parton}} - {p_T}^{\text{particle}}\right)$, and the parameters of the double-Gaussian are determined in a similar way as described in Sec. 5.4. The corrections for jets with cone-size $R = 0.4$ are shown in Fig. 5.6. The correction is about +18% for jets with absolute energy of $\sim$ 20 GeV,

and decreases asymptotically to about 2% for high-energy jets.

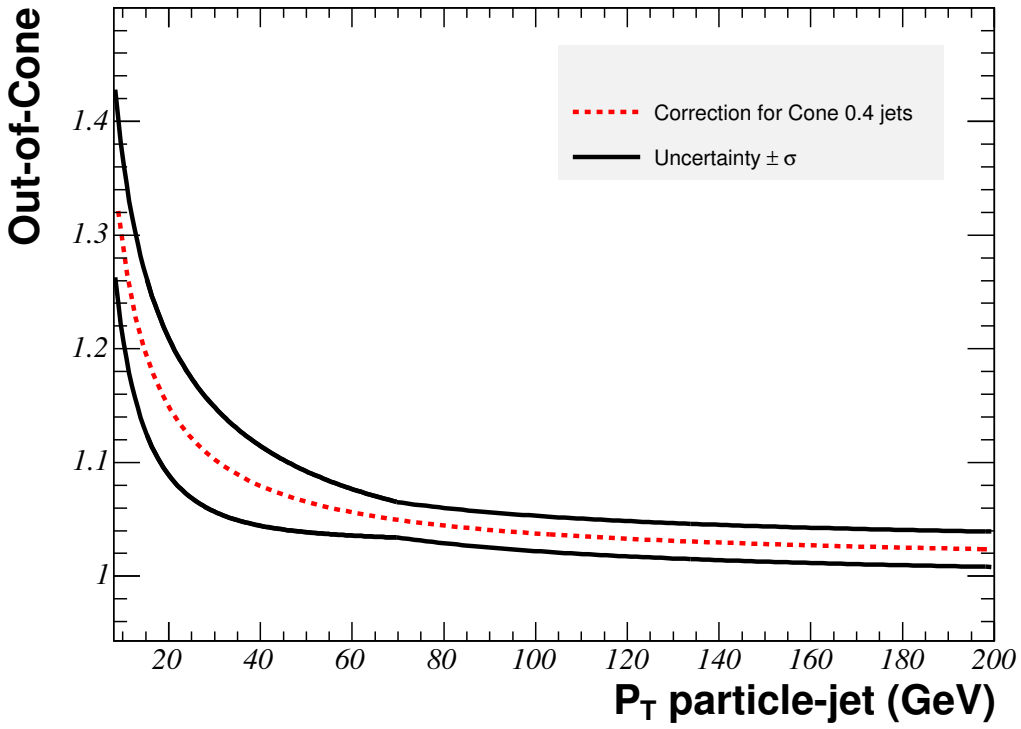

Figure 5.6: The out-of-cone energy correction as a function of p_T^{particle}, the p_T of the particle jet to which we apply the correction.

The largest uncertainty on this correction comes from modeling differences in the out-of-cone component of the correction between data and MC. These differences are determined in a photon-jet balancing sample, where a well-measured photon is compared to the measured energy of the jet. We measure the transverse energy in rings around the jet:

$$p_T(r_1 - r_2) = \sqrt{\left(\sum_{i=1}^{N} E_x^i\right)^2 + \left(\sum_{i=1}^{N} E_y^i\right)^2}$$

where N is the number of towers between a radius of r_1 and r_2 around the center of the jet. We compare the mean value of this quantity as a function of the corrected jet energies for $r_1 = 1.3$ and $r_2 = 0.4$ in data and MC, and see a difference of about 6% for low E_T jets that

decreases to about 2% for larger jet energies. Additionally, smaller uncertainties come from differences in the underlying event modeling in PYTHIA and HERWIG MC samples, and from measuring the small amounts of energy that fall outside a cone of $R = 1.3$, called splash-out.

5.6 Monte Carlo Quark/Gluon Jet Energy Calibration

The previous jet energy scale corrections are the standard energy corrections applied by CDF. Typically, analyses use jets corrected to the absolute energy level (internally called "Level 5") or to the parton level (internally called "Level 7"). However, as mentioned before, these energy corrections do not explicitly distinguish between the response of gluon and quark jets. The largest energy corrections, the absolute energy corrections and the out-of-cone corrections, are derived using PYTHIA dijet Monte Carlo simulations. Differences in the response of gluon and quark jets between MC and data may lead to differences in the measured energies of these objects, and these differences may not be covered by the existing systematic uncertainties on the jet energy scale.

It is possible to derive a correction for the response of quark and gluon jets in data and MC given two independent samples of jets, with different quark fractions, balanced against objects of known energy. We use events where a jet balances with a γ, which are rich in quark jets, and utilize the significant number of $Z \to \ell^+\ell^-$ + jet events now available, which are more rich in gluon jets. We construct the balance of the jet with these better measured reference objects:

$$K_{Z/\gamma} = (E_T{}^{\text{jet}}/p_T^{Z/\gamma}) - 1 \ . \tag{5.1}$$

For well-measured jets after the out-of-cone corrections, $K_{z/\gamma} = 0$. Rather than derive full

and separate jet energy scale corrections for quark and gluon jets in data and MC, we compare the balance in data and MC and derive an additional correction to be applied to MC jets, based upon whether they are matched to quarks or gluons.

In this section, we describe in detail the selection of both the Z-jet and γ-jet balancing samples, how well-balanced each sample is (*i.e.*, our determination of $K_{Z/\gamma}$), our method for extracting the balance of quark jets and gluon jets (K_q and K_g, respectively), and the final determination of the correction to apply to quark and gluon jets in the MC.

5.6.1 Z-Jet Balancing Selection

The datasets and event selection for Z-jet balancing sample closely follow that described in Sec. 4.2. These selection cuts are summarized in Tab. 5.1. We require two leptons consistent with being from the decay of a Z, where the two leptons come from the same lepton-type pairs used in the full analysis. Each lepton's p_T must be greater than 20 GeV/c, the reconstructed dilepton mass must be in a window centered around the mass of the Z boson, and we additionally require that the reconstructed Z boson p_T be greater than 10 GeV/c.

The jets used in these balancing studies are similar to the ones that are used in our search for $ZW/ZZ \to \ell\ell jj$. They have a cone size of R = 0.4, and are restricted to have $|\eta| < 2.4$ and an EM fraction > 0.9. Jets, unless otherwise stated, are corrected to Level 7 jet energies. In order to ensure the jet and Z boson are well-balanced, we allow only one jet in each event, eliminating events with additional energy clusters (electromagnetic or hadronic) above 3 GeV. We also require that the reconstructed Z boson and jet be back-to-back in the detector, with $\Delta\phi(Z, \text{jet}) > 2.8$ rad.

In Figs. 5.7-5.10 we show distributions of the balancing variable as defined in Eq. 5.1, $K_Z = {E_T}^{\text{jet}}/p_T^Z - 1$, in data and MC. These distributions are constructed in bins of jet E_T,

Z-jet Balancing Selection
$N_{\text{vtx}} > 0$
1st/2nd lepton $p_T > 20$ GeV/c
76 GeV/$c^2 < M_{\ell\ell} <$ 106 GeV/c^2
Z $p_T > 10$ GeV/c
$\not\!\!E_T < 20$ GeV
$N_{\text{jets}} > 3$ GeV = 1
$\Delta\phi$ between Z and jet > 2.8 rad

Table 5.1: Event selection requirements for the Z-jet balancing studies.

which allows us to derive a correction as a function of E_T^{jet}. Included in Figs. 5.7-5.10 is the fraction of the jets originating from quarks as determined by the MC samples: we see that the quark fraction is low ($\sim$ 35% for jets around 25 GeV), and increases with jet E_T. More about the quark fractions will be discussed in Sec. 5.6.3.

To describe the balancing in each bin, we fit around the most probable value of the distribution with a Gaussian, and use the mean and error on the mean derived from that fit as the value of $K_{Z/\gamma}$ for that jet E_T bin. The fitting range is chosen based on the quality of the fits, which are checked visually and quantitatively via the χ^2, and on robustness of the fitted mean. We tend to focus on fitting the distributions away from their long tails. It is because of these long tails that we do not simply use the mean of the balancing distribution, which are more dramatically affected by a small number of poorly measured jets. Furthermore, the mean and the median of the balancing distributions are both more affected by dijet backgrounds in the γ-jet balancing sample described in Sec. 5.6.2, and by trigger-sculpting from the high-E_T photon trigger used to select γ-jet events.

These Gaussian fits to the K_Z distributions in data and MC are also shown in Figs. 5.7-5.10. Fig. 5.11 shows the mean of these Gaussian fits as a function of the jet E_T. We see (1) that neither MC or data are particularly well-balanced across all jet E_T's, but more importantly (2) that the MC and data show poor agreement across jet E_T, indicating that after our energy corrections are applied, we still see large differences in the reconstruction

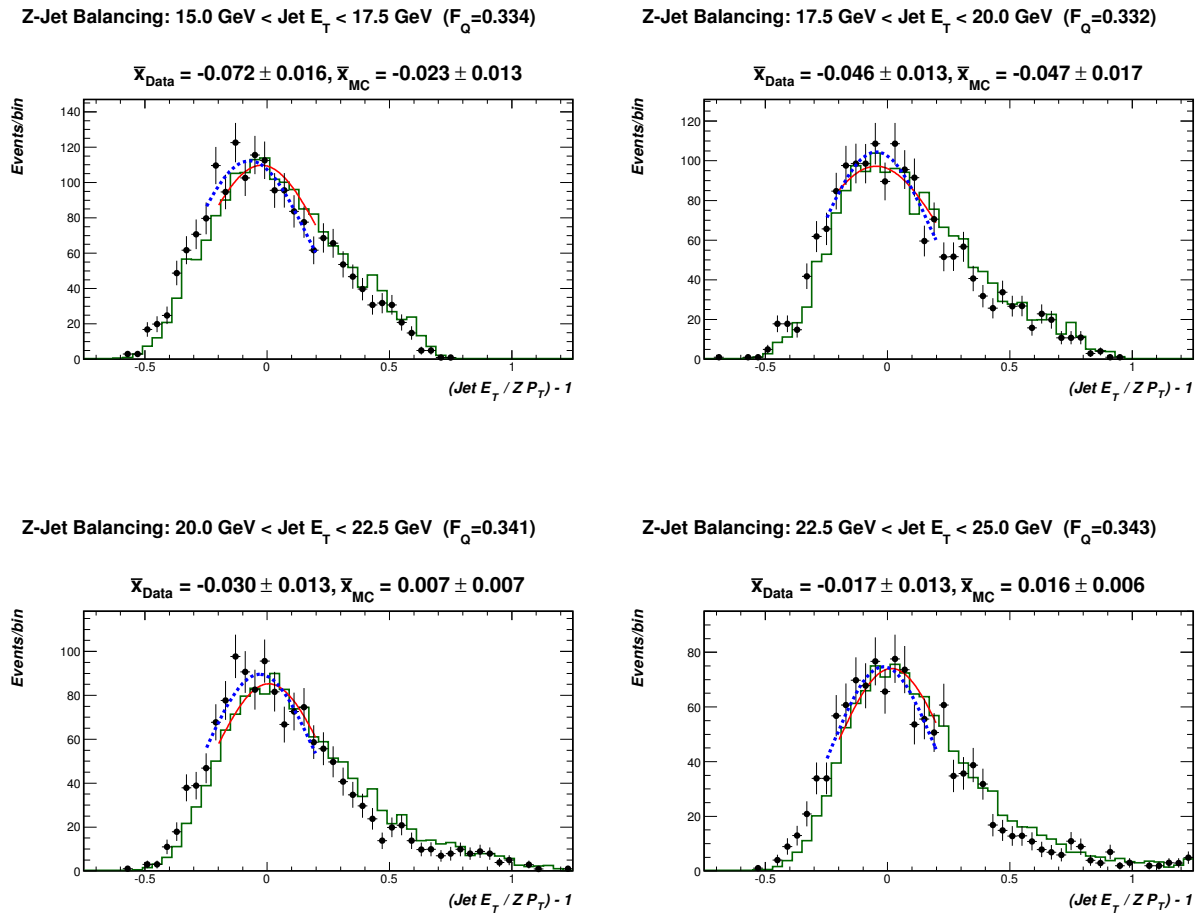

Figure 5.7: The balancing variable, K_Z, distribution in data (black) and MC (green) for E_T^{jet} between 15.0 - 17.5 GeV (upper left), 17.5 - 20.0 GeV (upper right), 20.0 - 22.5 GeV (lower left), and 22.5 - 25.0 GeV (lower right). Means and relative errors from Gaussian fits around the peaks are shown, along with the fraction of quark jets in MC.

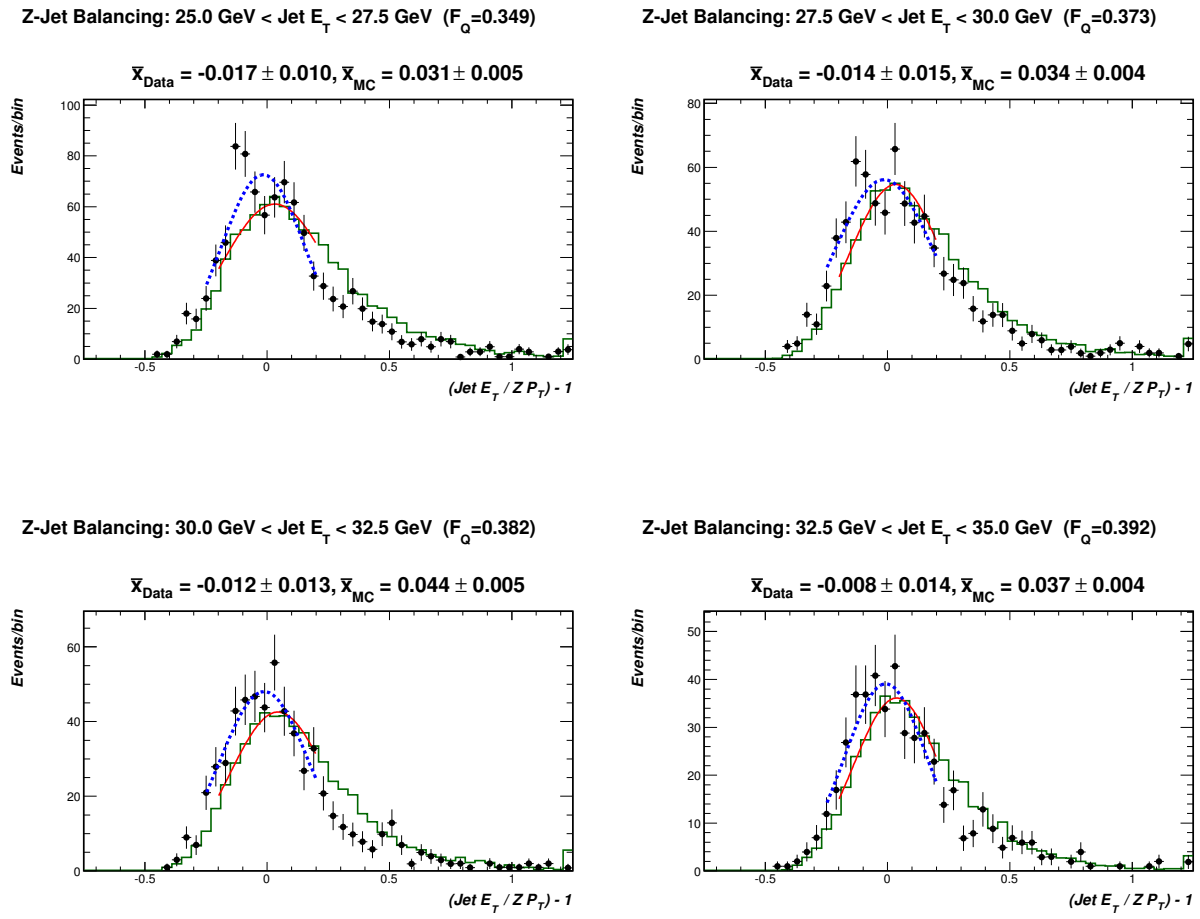

Figure 5.8: The balancing variable, K_Z, distribution in data (black) and MC (green) for E_T^{jet} between 25.0 - 27.5 GeV (upper left), 27.5 - 30.0 GeV (upper right), 30.0 - 32.5 GeV (lower left), and 32.5 - 35.0 GeV (lower right). Means and relative errors from Gaussian fits around the peaks are shown, along with the fraction of quark jets in MC.

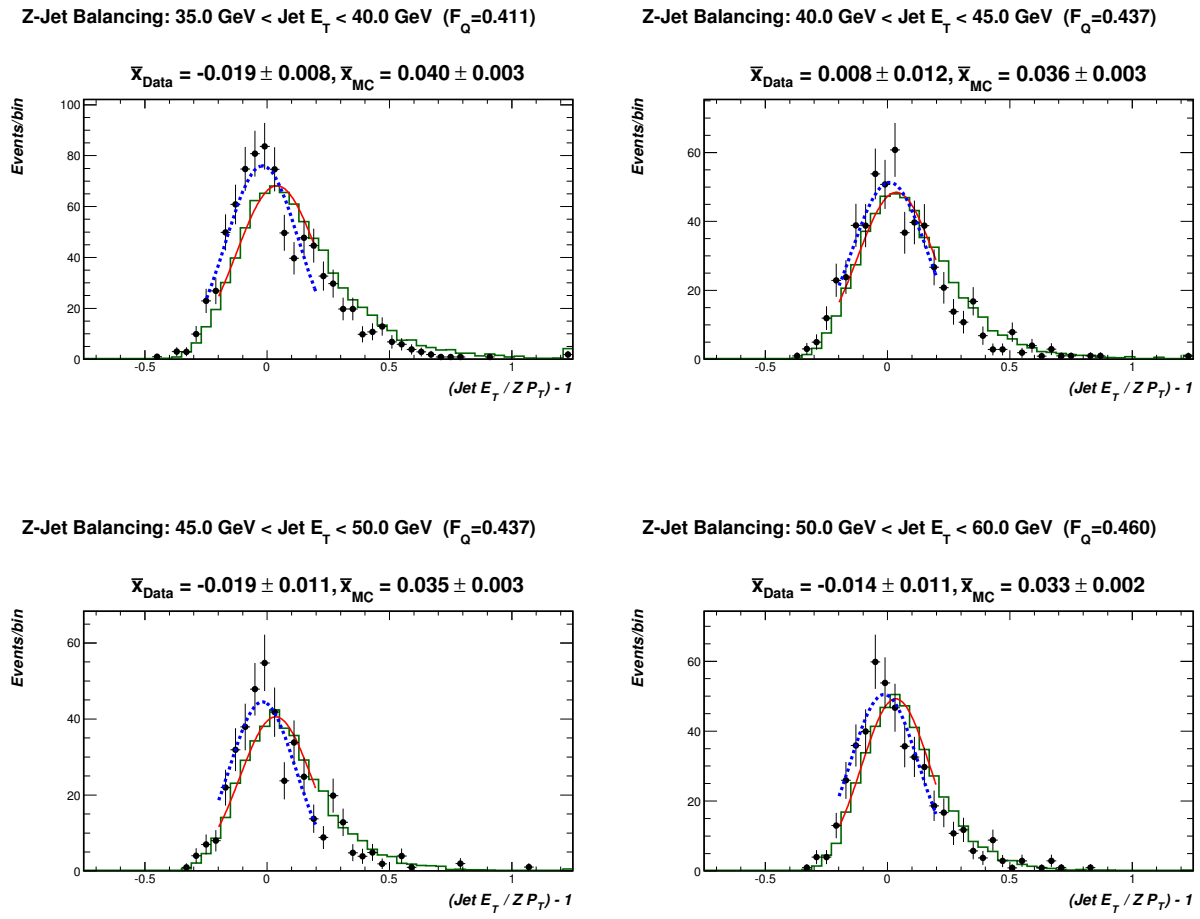

Figure 5.9: The balancing variable, K_Z, distribution in data (black) and MC (green) for E_T^{jet} between 35 - 40 GeV (upper left), 40 - 45 GeV (upper right), 45 - 50 GeV (lower left), and 50 - 60 GeV (lower right). Means and relative errors from Gaussian fits around the peaks are shown, along with the fraction of quark jets in MC.

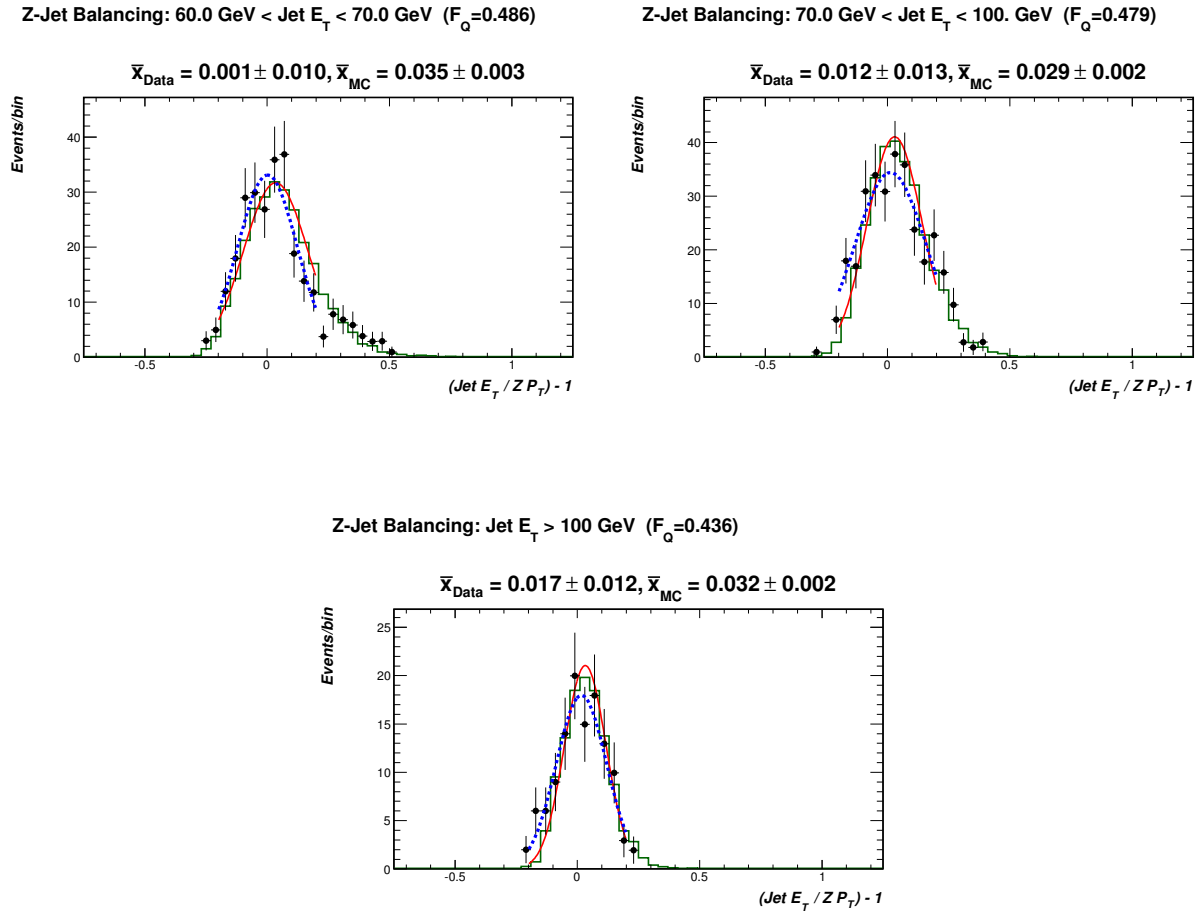

Figure 5.10: The balancing variable, K_Z, distribution in data (black) and MC (green) for E_T^{jet} between 60 - 70 GeV (upper left), 70 - 100 GeV (upper right), and greater than 100 GeV (bottom). Means and relative errors from Gaussian fits around the peaks are shown, along with the fraction of quark jets in MC.

of jets in data and MC.

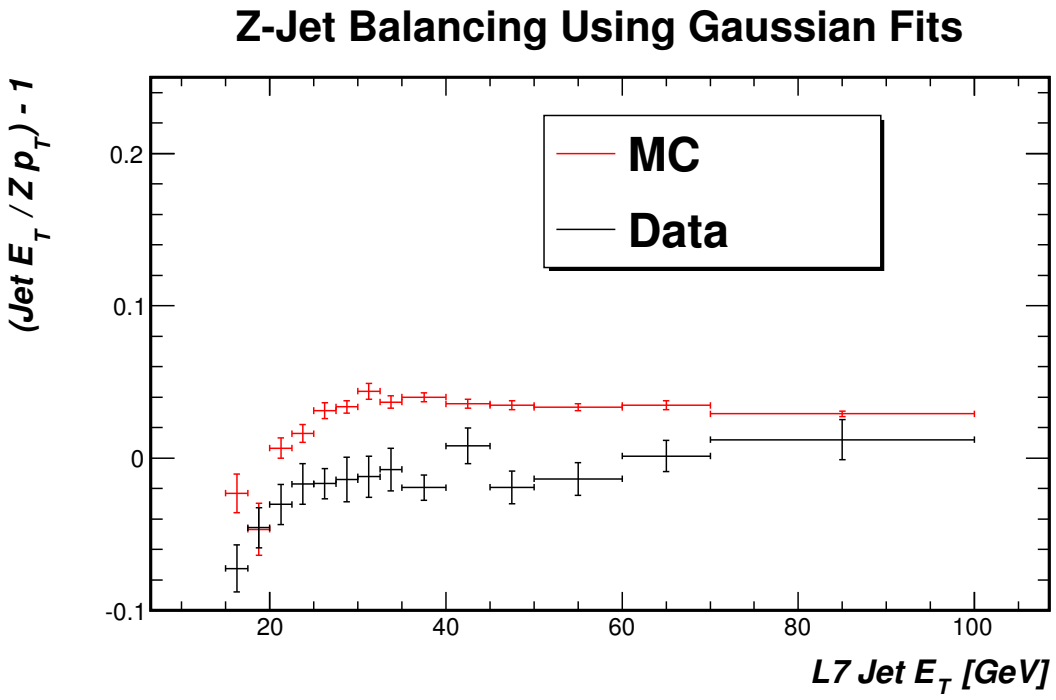

Figure 5.11: The balancing distribution, K_Z, in data (*black*) and MC (*red*) as a function of E_T^{jet}. The uncertainties are solely the uncertainty on the mean of a Gaussian fit to the balancing distributions in bins of E_T^{jet}. We see poor agreement in the Z-jet balancing, indicating errors in our reconstruction of jets in this sample.

5.6.2 γ-Jet Balancing Selection

For the γ-jet balancing sample, we use events collected with an isolated central photon trigger—requiring an isolated, 25 GeV photon—over the same data period as that of the high-p_T lepton samples. We compare this data to γ+ jet MC generated using PYTHIA, and also use PYTHIA MC samples to estimate the contributions from dijet production that contaminate our γ-jet balancing sample. Both the γ+ jet and dijet MC samples are generated while varying a minimum requirement on the p_T of the generated partons, $\hat{p_T}$. For the γ+ jet MC, that cut ranges from as low as 13 GeV/c to a sample with $\hat{p_T} > 70$ GeV/c.

In the dijet MC, that range is from 18 GeV/c to a high-$\hat{p_T}$ sample with $\hat{p_T} > 230$ GeV/c. These samples are stitched together by placing cuts on the generator-level parton p_T. This ensures that we have a large number of MC events across all jet E_T ranges.

The selection requirements for our γ-jet balancing sample are summarized in Tab. 5.2. We closely follow the selection requirements used in previous γ-jet balancing studies [48]. We require a high-quality central photon following the selection requirements of [59]. In order to avoid trigger biases, we require $E_T^\gamma > 27$ GeV and $0.2 \leq |\eta_\gamma| \leq 0.6$ in both data and MC. To decrease the contribution from dijet production, where a jet mimics our photon selection, we require the energy in the calorimeter and momentum in the tracking system contained within a cone of $R = 0.4$ around the photon to be less than 1 GeV and 2 GeV/c, respectively. As in the Z-jet balancing sample, we require events have one and only one jet with (uncorrected) $E_T > 3$ GeV within $|\eta| \leq 2.4$, and demand the $\Delta\phi$ between the jet and photon be > 3.0 radians. We further reduce contamination of this sample by vetoing events with large pileup (more than one reconstructed interaction point), and by removing events with $\not{E}_T/E_T^\gamma > 0.8$, which likely contain activity from cosmic rays.

γ-jet Balancing Selection
$E_T^\gamma > 27.5$ GeV
$0.2 < \|\eta_\gamma\| < 0.6$
Calorimeter Isolation ($R = 0.4$) < 1 GeV
Track Isolation ($R = 0.4$) < 2 GeV/c
$N_{vert} = 1$
$\not{E}_T/E_T^\gamma < 0.8$
$N_{jets} > 3$ GeV $= 1$
$\Delta\phi$ between Z and jet > 3.0 rad

Table 5.2: Event selection requirements for the γ-jet balancing studies.

The distribution of the balancing variable K_γ (see Eq. 5.1) is shown in Figs. 5.12-5.14 in data and MC, in the same jet E_T bins as shown in Sec. 5.6.1, except with $E_T^\gamma > 27.5$ GeV. We use the same procedure of fitting the distribution with a Gaussian around the peak to

extract a most probable balancing variable, and show the distribution of K_γ as a function of jet E_T in Fig. 5.15. Here, while the jets do not perfectly balance the photon, the agreement between data and MC is good: our MC simulation models the jets from our γ-jet sample well.

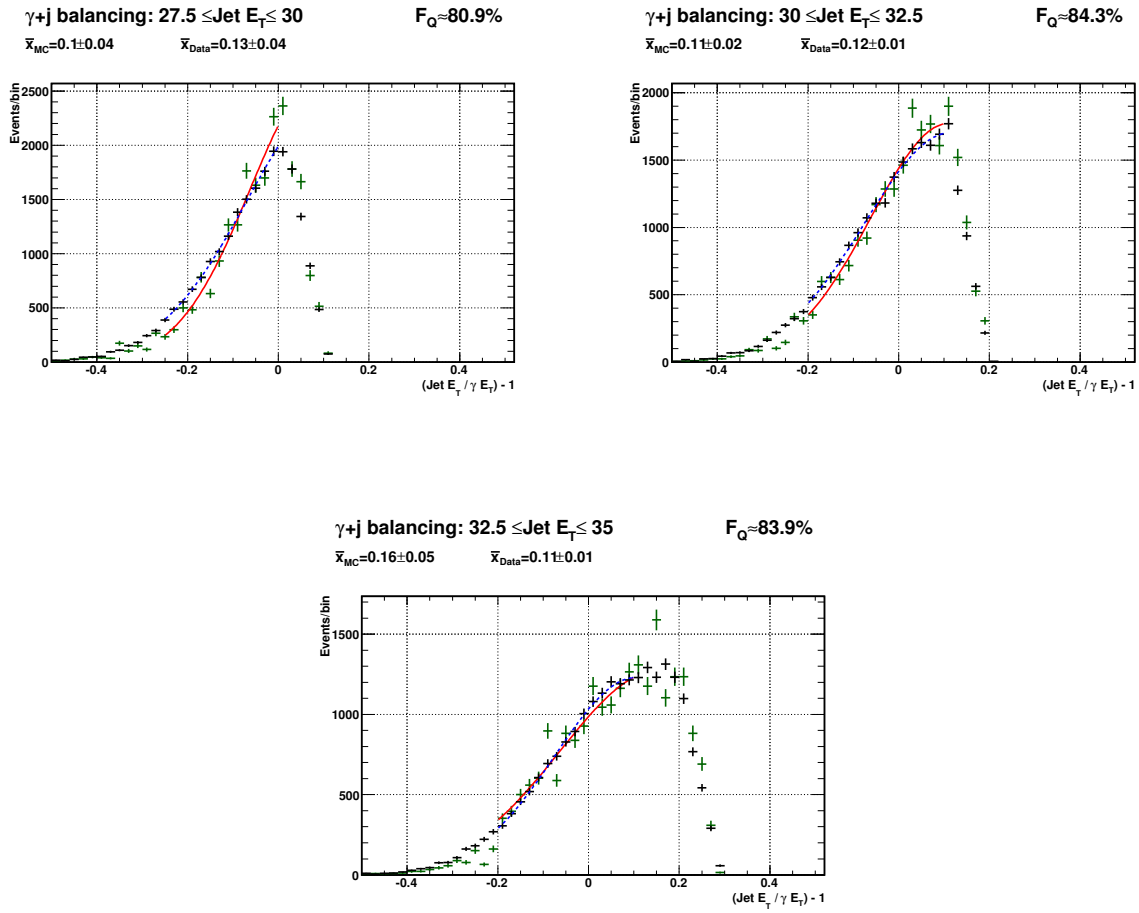

Figure 5.12: *K_γ distribution in data (black) and MC (green) for E_T^{jet} / GeV bin 27.5-30 (upper left), 30-32.5 (upper right), 32.5-35 (lower). Means and relative errors from the gaussian fits are also shown.*

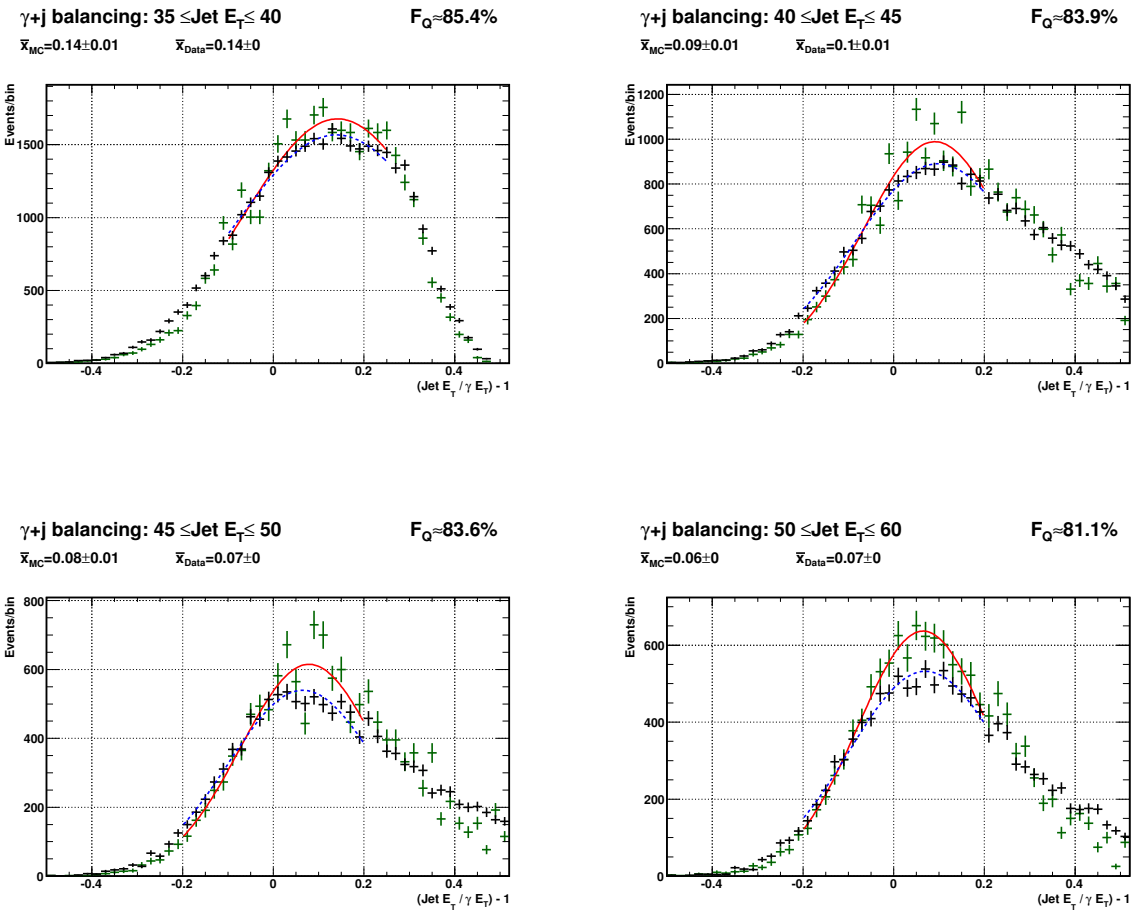

Figure 5.13: *K_γ distribution in data (black) and MC (green) for E_T^{jet} / GeV bin 35-40 (upper left), 40-45 (upper right), 45-50 (lower left), 50-60 (lower right). Means and relative errors from the gaussian fits are also shown.*

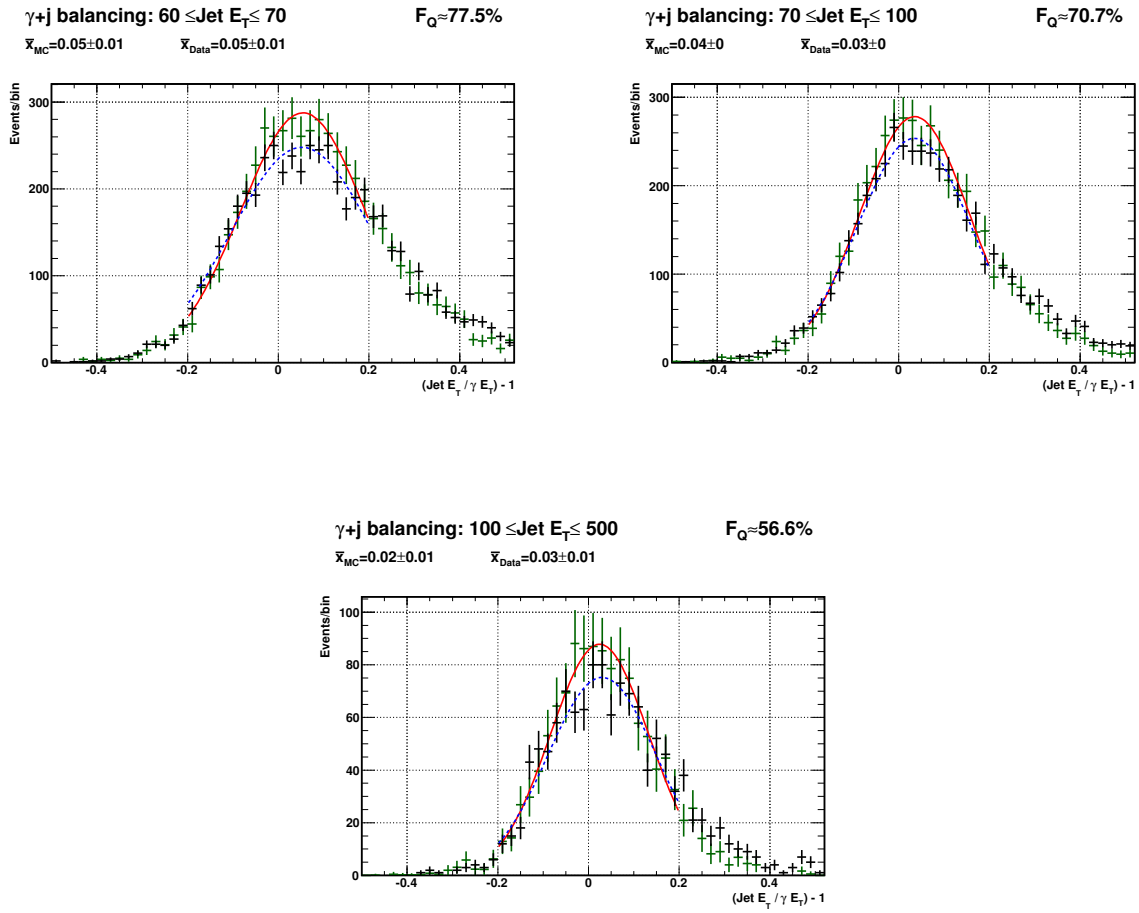

Figure 5.14: *K_γ distribution in data (black) and MC (green) for E_T^{jet}/GeV bin 60-70 (upper left), 70-100 (upper right), 100-500 (lower). Means and relative errors from the gaussian fits are also shown.*

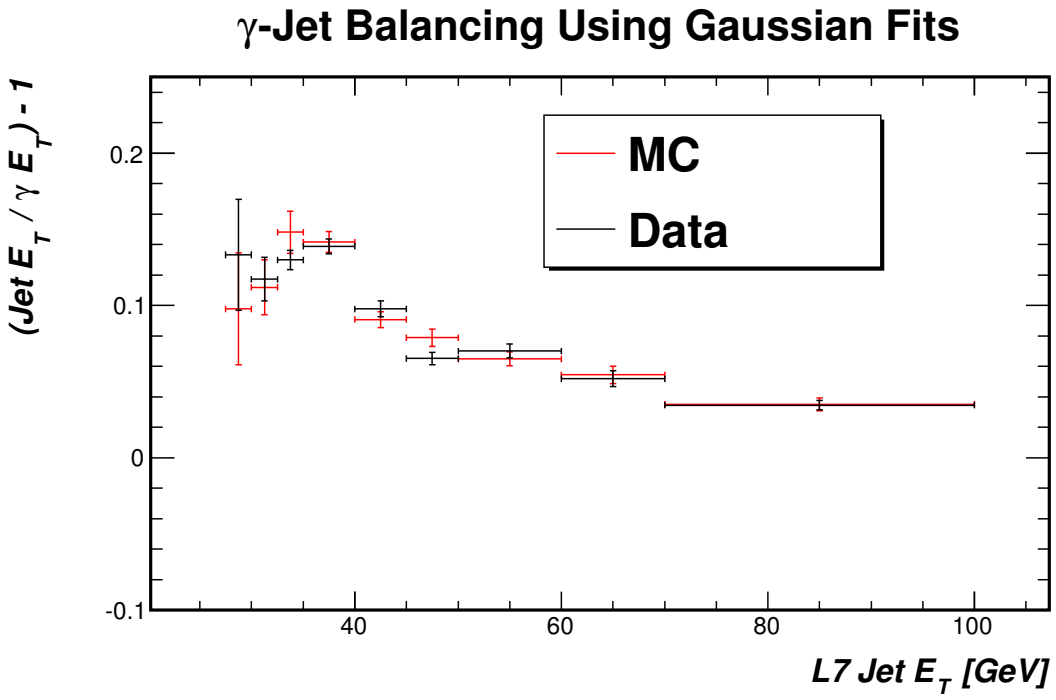

Figure 5.15: The balancing distribution, K_γ, in data (*black*) and MC (*red*) as a function of E_T^{jet}. The uncertainties are solely the uncertainty on the mean of a Gaussian fit to the balancing distributions in bins of E_T^{jet}. We see better agreement in the γ-jet balancing than in the Z-jet balancing (see Fig. 5.11); the jets here are well-modeled by MC, and we interpret this difference as being due to the difference in the quark/gluon composition of each sample.

5.6.3 Separate Corrections for Quark and Gluon Jets

Comparing Figs. 5.11 and 5.15, there is a clear difference in our modeling of jets from these two samples: our MC models jets from the γ+ jet selection well, but not those from the Z+ jet selection. While the systematic uncertainties for the jet energy correction to the parton level are on the order of $\sim$ 5% for jets with E_T = 30 GeV, and thus may (somewhat) cover the difference between data and MC jets in the Z-jet balancing sample, there is not a corresponding discrepancy in the γ-jet sample that should be there in the case of a correlated systematic difference. The difference in the accuracy of the MC modeling in these two selections indicates an additional correction is necessary that will fix the modeling in the Z-jet sample, while not greatly affecting the modeling in the γ-jet sample.

One of the biggest differences in the jets in these two samples is their parton progenitor: jets from γ+ jet events are largely quark jets, while those from Z+ jet events have a much higher proportion originating from gluons. Gluon jets tend to be more spatially spread than quark jets, with a higher particle multiplicity and lower particle p_T spectrum for a given jet energy. Thus, quark and gluon jets will likely be reconstructed differently in the calorimeter, but the lack of independent corrections for quark and gluon jets means these differences are averaged over in our corrections. Furthermore, the largest energy corrections are the absolute (L5) and out-of-cone (L7) corrections, which are completely determined from MC simulation—the only point at which jets from data contributes here is in the determination of the systematic uncertainty for the L7 correction, but that is in a quark-dominated γ+jet sample. A difference in the modeling of gluon jets between data and MC could lead to the the discrepancy seen in Z+ jet events while reinforcing the agreement in γ+ jet events.

To work towards deriving separate corrections for quark and gluon jet energies in MC, we use our Z-jet and γ-jet balancing samples in the following way. K_Z and K_γ are the

necessary corrections to jet energies in the Z-jet and γ-jet balancing samples, as defined in Eq. 5.1. Each correction is a weighted average of separate corrections for quark and gluon jets: K_q and K_g, respectively. If $F_X^{q/g}$ is the quark/gluon fraction in sample X, then we can write:

$$K_Z = F_Z^q K_q + F_Z^g K_g = F_Z^q K_q + (1 - F_Z^q) K_g \tag{5.2}$$

$$K_\gamma = F_\gamma^q K_q + F_\gamma^g K_g = F_\gamma^q K_q + (1 - F_\gamma^q) K_g \ . \tag{5.3}$$

These equations can be rewritten, solving for K_q and K_g:

$$K_q = \frac{1}{F_\gamma^q - F_Z^q}[(1 - F_Z^q) K_\gamma - (1 - F_\gamma^q) K_Z] \tag{5.4}$$

$$K_g = \frac{1}{F_\gamma^q - F_Z^q}[F_\gamma^q K_Z - F_Z^q K_\gamma] \ . \tag{5.5}$$

These equations may be written separately for data and MC (thus with distinct $K_X{}^{\text{data}}$ and $K_X{}^{\text{MC}}$), and may include a dependence on the energy of the jet ($F_X^q \to F_X^q(E_T{}^{\text{jet}})$ and $K_X \to K_X(E_T{}^{\text{jet}})$). In order to solve for K_q and K_g, we need to know $K_{Z/\gamma}$, for which we use the balancing derived in Secs. 5.6.1 and 5.6.2.

This procedure for determining K_q and K_g also relies on the fact that the fraction of quark jets in our two samples is different and known to some degree of precision. The determination of $F_{Z/\gamma}^q$ in MC is fairly trivial, as we may just match jets to their originating parton. We match jets by looping through the particles in the event record, and locating the highest p_T parton located inside the cone of the jet. In our γ-jet balancing sample, we find that the quark fraction is about 85% at $E_T{}^{\text{jet}} \sim 30$ GeV, and drops to about 71% at $E_T{}^{\text{jet}} \sim 70$ GeV. In the Z-jet balancing sample, these fractions are $\sim 38\%$ and $\sim 49\%$ at the same $E_T{}^{\text{jet}}$ points.

However, in data we do not know the quark/gluon fraction *a priori*. Furthermore, we cannot assume the $F_q(E_T{}^{\text{jet}})$ from MC, as the jet E_T distribution in data and MC may be dif-

ferent (which is what motivates our work towards developing a correction). Instead, since the Z p_T and γ E_T are well-measured, we can determine the quark fraction as a function of these values, $F_q(p_T^{Z/\gamma})$, in MC, and assume the data behaves similarly. The quark fractions, $F_q(p_T^{Z/\gamma})$, as determined from MC, are shown in Fig. 5.16. We parameterize $F^q_{Z/\gamma}$ as a function of $p_T^{Z/\gamma}$,

$$F^q_{Z/\gamma}{}^{\mathrm{MC}}(p_T) = a + e^{bp_t+c} \ ,$$

and determine the $F^q_{Z/\gamma}{}^{\mathrm{data}}$ in each jet E_T bin of the data based on $p_T^{Z/\gamma}$ distribution in the data.

We establish an uncertainty on F_q by looking at the difference in the distributions of the jet QG value in data and MC. The jet QG value, described in further detail in Chap. 7, is from artificial neural-network discriminant that examines the shape of a jet and assigns a score based on how quark-like the jet appears. The discriminant is sensitive to differences in the quark/gluon fraction of a given sample. However, the discriminant is also sensitive to other variations in the shapes of jets, and so we use it only for establishing a conservative uncertainty on the quark fraction obtained from MC. We fit the distribution of jet QG values in data with quark and gluon templates from MC to extract a new F_q, and take the difference between this value and the one obtained via matching in MC as a systematic uncertainty. Typically, these uncertainties are about 10%, which we use as an uncertainty on the F_q values in both samples.

Using Eqs. 5.4-5.5, we construct distributions of K_q and K_g as a function of the jet E_T, shown in Fig. 5.17. The uncertainties in these values are determined by assuming Gaussian error propagation of the errors from K_Z, K_γ, F_q^Z, and F_q^γ. The distribution of $K_q(E_T{}^{\mathrm{jet}})$ is very similar to that of $K_\gamma(E_T{}^{\mathrm{jet}})$ (see Fig. 5.15), since the γ-jet balancing sample is very quark-rich. Meanwhile, the distribution of $K_g(E_T{}^{\mathrm{jet}})$ looks more akin to $K_Z(E_T{}^{\mathrm{jet}})$, given the higher gluon fraction in the Z-jet sample. Furthermore, we see good agreement

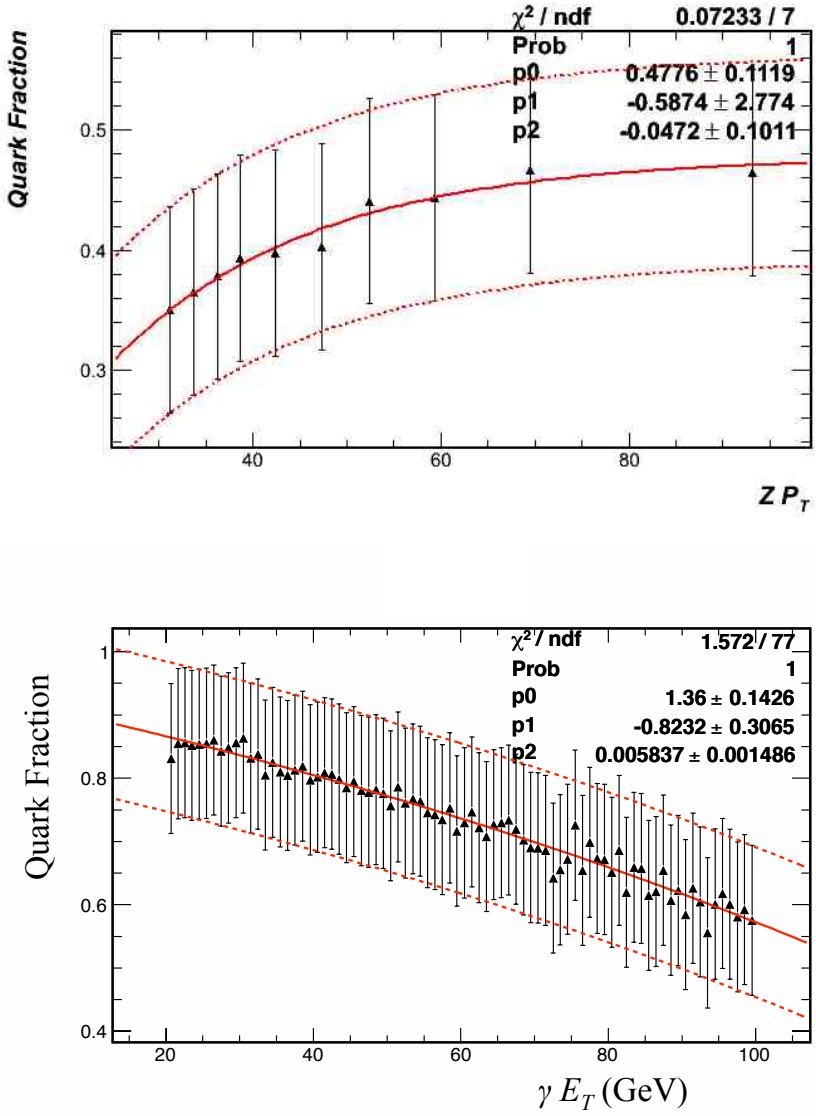

Figure 5.16: The fraction of quark jets in Z+jet events ($F_q(p_T^Z)$, top) and in γ+jet events ($F_q(E_T^\gamma)$, bottom). The uncertainties are derived from fits to the jet QG value observed in data, and thus are rather correlated bin-to-bin. The fractions are fit in ROOT using an exponential plus a constant: $\mathrm{p0} + e^{\mathrm{p1}+\mathrm{p2}\cdot x}$. The parameters of the functional fit are listed in the plots.

between data and MC in K_q, but worse agreement in data and MC in K_g. The fact that data appears consistently lower than the MC in the $K_g(E_T^{\text{jet}})$ distribution indicates that MC is systematically overestimating gluon jet energies, relative to the data.

Now with the distributions of K_q and K_g, we determine the corrections that need to be applied to MC jets in order to best match the energy scale of the data. These MC corrections are defined as $({K_q}^{\text{Data}}+1)/({K_q}^{\text{MC}}+1)$ for quark jets, and $({K_g}^{\text{Data}}+1)/({K_g}^{\text{MC}}+1)$ for gluon jets, and are shown in Fig. 5.18. Due to the photon trigger used to select the γ-jet balancing sample, we do not have reliable balancing information for jets below 27.5 GeV in that sample, limiting the full range over which we may derive corrections. Since we are interested in jets down to energies around 20 GeV, we extrapolate to lower jet energies the quark jet energy correction derived for jets with $E_T \geq 27.5$ GeV, and use the Z-jet balancing sample to extract a gluon correction assuming this extrapolated quark correction.

Both the quark and gluon corrections appear flat in jet energy for jets with $E_T \geq$ 15 GeV, and so we fit them to a constant. We find that to better match the data, quark jet energies in MC should be increased by $\sim$ 1.4%, while gluon jet energies should be decreased by $\sim$ 7.9%. This follows what we expected based on the Z-jet and γ-jet balancing samples: MC models quark jets well, and so the correction for quark jets is small (and consistent with no correction necessary, as discussed in Sec. 5.6.4); however, MC does not model gluon jet energies well, overestimating their energies by a significant amount and so requires a significant shift downward in the gluon jet energies.

5.6.4 Uncertainties on MC Jet Energy Corrections

We consider the following sources of error on the corrections presented in Sec. 5.6.3. Because the corrections shift the energy response in MC to better match data, the quark

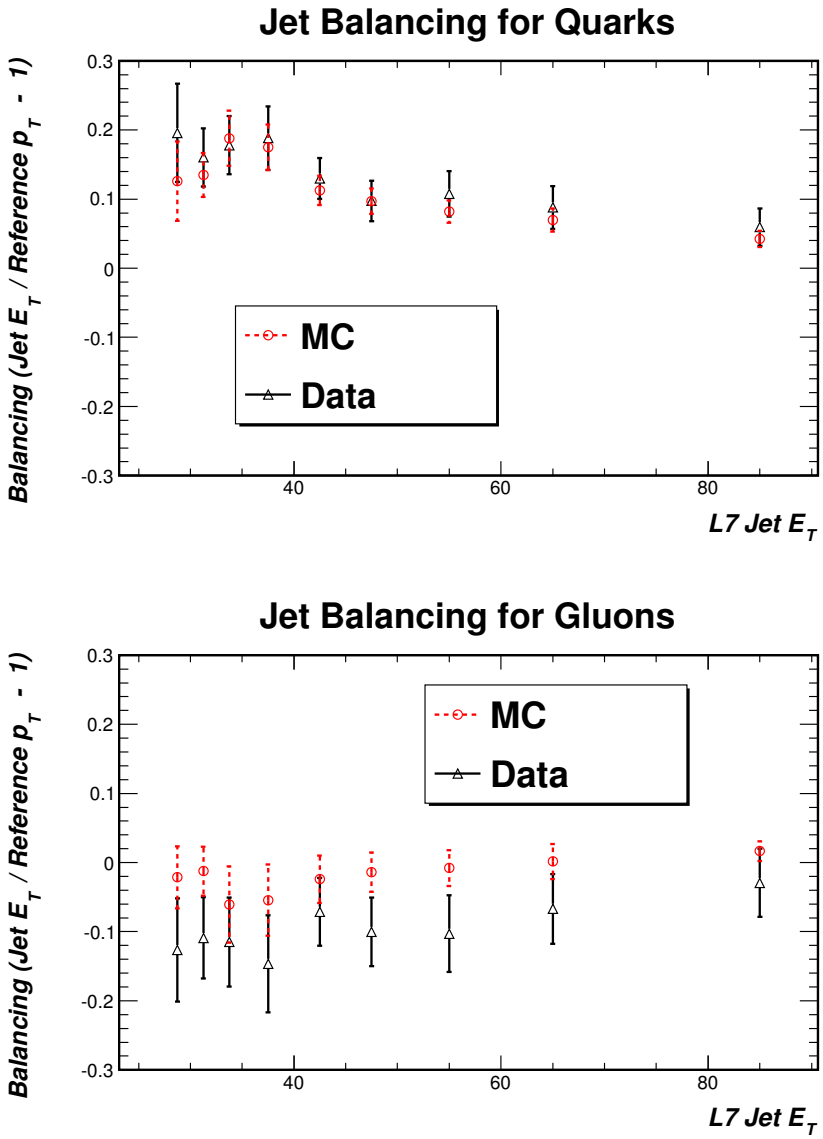

Figure 5.17: The derived balancing variable for quark jets, K_q, (*top*) and gluon jets, K_g, (*bottom*) in data (*black*) and MC (*red*) as a function of E_T^{jet}. The uncertainties on each point are from the uncertainties from the mean of the Gaussian fit and the uncertainties on the quark fractions, added in quadrature. We see better agreement between data and MC in the energy scale of quark jets than that of gluon jets, following from the behavior seen in Figs. 5.11 and 5.15.

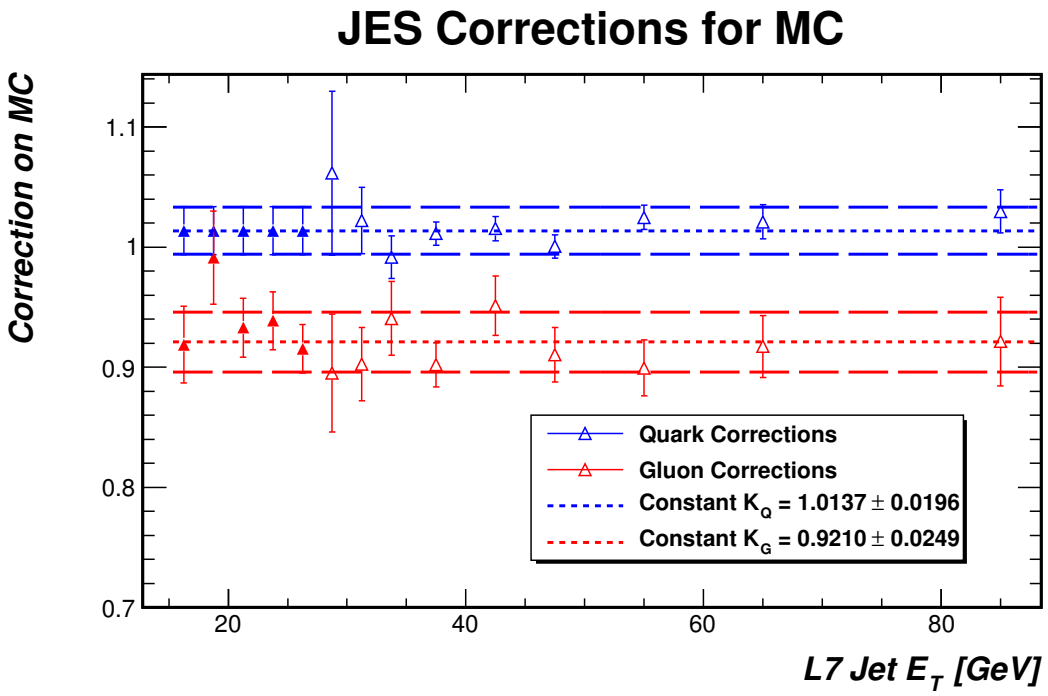

Figure 5.18: The derived correction for MC quark jets (*blue*) and gluon jets (*red*) as a function of E_T^{jet}. The open triangles represent corrections derived using both γ-jet and Z-jet balancing samples, while the filled triangles represent the assumed flat correction for quarks and the corresponding correction for gluons calculated from the Z-jet balancing sample alone. The error bars represent only the statistical uncertainty on the balancing variable. The short dashed lines are the fits of the correction to a constant across jet E_T, and the long dashed lines represent the error bands represent just the statistical uncertainty on the fit, further described in Sec. 5.6.4.

jet and gluon jet energy correction uncertainties are anti-correlated: if the quark jet energy correction goes up, the gluon jet energy correction must go down in order to further compensate for that shift, and *vice versa*.

- *Fit/Statistical Uncertainty*: We use the standard deviation of the necessary MC corrections of each jet E_T bin to capture the spread of the MC corrections around the assumed flat correction function. This is an uncertainty of $\pm$2.0% for quark jet energies, and $\mp$2.5% for gluon jet energies.

- F_q^Z: We vary the the quark fraction of the Z-jet sample by $\pm$10% (absolute, as described in Sec. 5.6.3), and recalculate the corrections for quark and gluon jets. This translates to an uncertainty of $\pm$0.6% for quark jet energies, and $\mp$2.1% for gluon jet energies.

- F_q^γ: Similarly, we vary the the quark fraction of the γ-jet sample by $\pm$10%. This translates to an uncertainty of $\pm$1.8% for quark jet energies, and $\mp$2.7% for gluon jet energies.

- *Low E_T Extrapolation*: We check the dependence of the gluon jet energy corrections on the assumed quark jet corrections for low E_T jets by moving the quark jet E_T for these jets by $\pm$2%, based on the uncertainty in the jet energy scale measured *in situ* in top mass measurements at CDF [60]. We see a small change in the fit gluon energy corrections: $\mp$0.4% of the jet energy.

- *Number of Interaction Vertices Dependence*: The γ-jet sample has a cut on the number of reconstructed interaction vertices in order to reduce contamination from pileup. The Z-jet sample does not place such a cut, in order to retain as many events as possible. We check the effect this cut has by checking for any shift in the corrections when the cut is placed on the Z-jet sample. We see a change to the

quark jet energies of $\pm 0.2\%$, and the gluon jet energies of $\mp 1.2\%$.

The uncertainties are summarized in Tab. 5.3. The uncertainties are similar in magnitude to the current energy scale uncertainties [48]. We replace the previously existing systematic uncertainties on the jet energy scale with these newly derived ones – rather than include the new uncertainties alongside the old ones – for two reasons. First, the bulk of the previously existing systematic uncertainties on the jet energy scale reflect possible differences in data and MC, which we are now accounting for in an improved way that depends on the originating parton with the new corrections. Second, our cross section measurement is not sensitive to changes in the real-value energy scale, but instead to differences between data and MC. Therefore, the total systematic uncertainty in Tab. 5.3 should represent the full necessary uncertainties on the jet energy scale.

		Quark jets	Gluon jets
JES Correction		1.014	0.921
Uncertainty	Fit/Statistics	0.020	0.025
	F_q^Z	0.006	0.021
	F_q^γ	0.018	0.027
	Low E_T Extrapolation		0.004
	N_{vert} difference	0.002	0.012
	Total	± 0.027	∓ 0.044

Table 5.3: Summary of the additional jet energy corrections applied to MC jets, and the uncertainty on those corrections. The uncertainties for the quark jet and gluon jet energy corrections are anti-correlated, as they must work in concert to match the balancing distributions in data.

Fig. 5.19 shows the necessary corrections for quark and gluon jet energies with the full set of uncertainties summarized in Tab. 5.3. We find that quark jets need a correction that increases the jet energy after L7 corrections of $+1.4\% \pm 2.7\%$. Gluon jets in MC must be corrected downwards to better agree with data, with a correction to the jet energy of $-7.9\% \mp 4.4\%$.

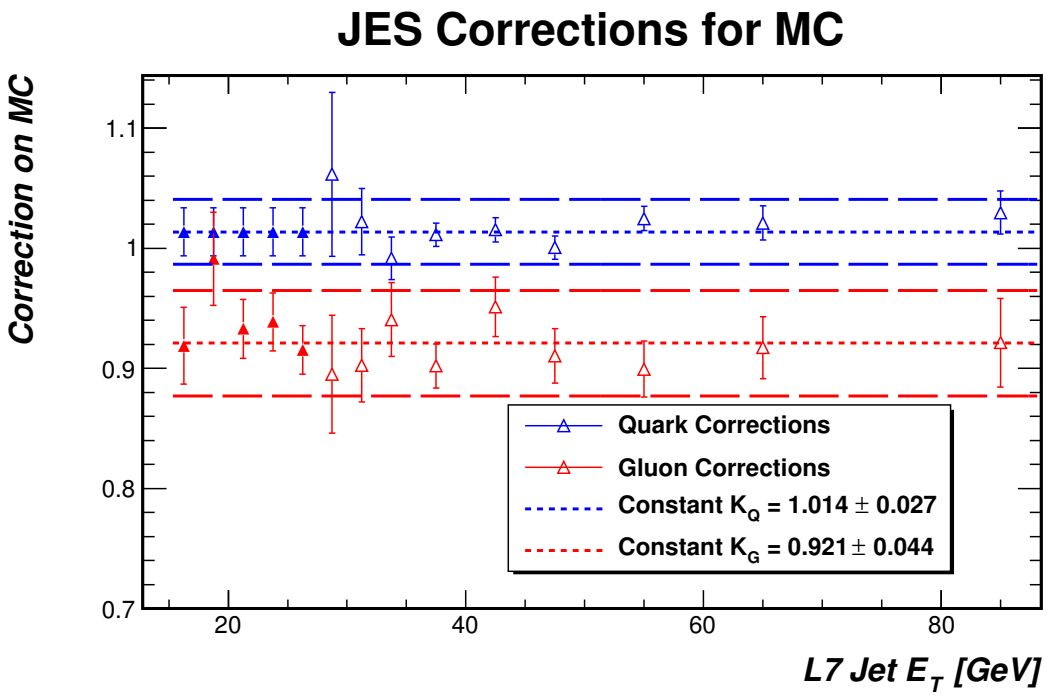

Figure 5.19: The derived correction for MC quark jets (*blue*) and gluon jets (*red*) as a function of E_T^{jet}. The open triangles represent corrections derived using both γ-jet and Z-jet balancing samples, while the filled triangles represent the assumed flat correction for quarks and the corresponding correction for gluons calculated from the Z-jet balancing sample alone. The error bars represent only the statistical uncertainty on the balancing variable. The short dashed lines are the fits of the correction to a constant across jet E_T, and the long dashed lines represent the error bands represent the full uncertainty, described in Sec. 5.6.4.

CHAPTER 6

NEURAL-NETWORK-BASED b JET IDENTIFICATION ALGORITHM

Identifying jets originating from b quarks is an important part of many analyses at high-energy particle colliders, including searches for the Higgs boson and measurements of top-quark properties. We are interested in b-tagging in this analysis due to the increased sensitivity we may have to $ZZ \to \ell^+\ell^- b\bar{b}$ decays. B hadrons have particular qualities that help distinguish b jets from other light flavor quark or gluon jets: they typically have high momentum and long lifetimes, resulting in a large distance between the interaction point and the decay vertex of the B hadron. Additionally, a significant fraction ($\approx$20%) of B hadrons decay in semi-leptonic modes, producing leptons with low p_T inside the cone of the jets. These qualities are key to distinguishing b-quark jets from other types of jets.

Here we describe an artificial neural-network-based b jet tagger that we employ in our ZW/ZZ search. The tagger is described in more detail in [61]. The tagger is unique in its emphasis on individual tracks, and in its ability to evaluate jets with only a single track. The output of this ANN, which we call the jet $b$ness, is designed to identify jets containing a B-hadron decay. To characterize the tagger's performance, the efficiency and mistag rate are obtained as a function of a cut on the jet $b$ness in $Z+1$ jet (rich in light flavor jets) and $t\bar{t}$ (rich in b jets) candidate samples. This choice differs from many previous evaluations of performance using generic dijet samples, but the large data sample accumulated at the Tevatron allows us to use the more pure top quark samples for b-tagging efficiency studies, and the momentum spectrum of b quarks in top pair production is better matched to diboson searches than the relatively soft quark momentum spectrum found in generic dijet samples.

6.1 Introduction

Identifying jets with B hadron decays can be accomplished through careful consideration of the tracks inside the jet. The lifetime of the most common B hadrons produced in $p\bar{p}$ collisions – *e.g.* the B^0, $B^{\pm}$, and B_s^0 mesons, and the Λ_b^0 baryon – is approximately 1.6 ps [18]. For relativistic B hadrons with $p \approx 20$ GeV/c, this translates to a decay length of about 2 mm, a distance resolvable with the CDF tracking system. Therefore, using a combination of the COT and silicon detectors, it is possible to identify B hadron decays by searching for individual tracks displaced from the primary interaction vertex, and by searching for multiple tracks coming from a displaced secondary vertex.

There are properties beyond the B hadrons' lifetimes that aid in b-jet identification. The decay products of B hadrons will generally form a larger invariant mass than the decays products of other hadrons, since the b quark (and thus its corresponding hadrons) has a larger mass than other quarks. In addition, when produced in jets, B hadrons tend to carry a large fraction of the jet energy, giving them and their decay products a large boost that makes these particles more energetic and collimated within the jet cone. Finally, B-hadron decays are relatively rich in semi-leptonic decays, with roughly 20% of decays yielding an electron or muon [18].

The above properties have formed the basis for previous b-jet tagging methods at CDF. The most commonly used tagger is SecVtx [62], a secondary vertex tagger which fits for secondary vertices using only significantly displaced tracks. The tagger determines the significance of the two-dimensional decay length in the r-ϕ plane – the L_{xy} – and uses this to select b-jet candidates. The jet probability [63] tagger, another b-tagging algorithm, does not look for a secondary vertex, but instead compares the distribution of the "impact parameter significance" (d_0/σ_{d_0}) for tracks inside a jet to the expected distribution from light jets, and assigns a probability a light-flavor jet would have tracks as or more

displaced. Additionally, soft-lepton taggers [64] exist that look solely for the existence of a soft lepton from a semi-leptonic *B* hadron decay inside the jet. While the low branching ratio and difficulty of identifying soft electrons and taus limits this approach, keying on soft muons allows a search for *b* jets that does not rely on displaced tracks.

Beyond these more conventional approaches to *b*-jet tagging, a number of taggers make use of artificial neural networks (ANNs, or sometimes simply NNs). An ANN is a simulated set of interconnected neurons, where each neuron produces a numerical response to a given set of input signals [65]. The neurons may be arranged in a way to take a set of input variables and produce a single-value output variable. The response of the neurons may be chosen so that, as a whole, the ANN acts as a discriminant, assigning a high ANN score to a signal that typically produces a certain set of inputs, and assigns a low score to backgrounds that behave differently. A number of parameters may be chosen in setting up an ANN: the input variables used, the number of neurons to use in the network, and into how many layers they should be arranged, and the type of response for each neuron. Fig. 6.1 shows an example of how the neurons in an ANN may be set up to produce a single-valued ANN score given a set of inputs. ANNs provide the ability to consider many variables independently, but can also exploit correlations in many variables and identify differences in how "signal" and "background" variables may be correlated.

Because ANNs can use as many discriminants as is computationally feasible, the power of ANNs can exceed that of more conventional taggers. A number of ANN-based taggers have been used in CDF analyses. One, the "KIT flavor separator" [66], is not so much a tagger in its own right, but supplements the SecVtx tagger by identifying SecVtx-tagged jets that more likely originate from a *b* quark than from either a *c* quark or from other partons. The "Roma tagger" [67, 68] uses a vertexing algorithm that can search for multiple vertices inside a jet, like those seen in cascade *B* hadron decays. It uses three types of ANNs that make use of information from the SecVtx, JetProb, and soft-lepton

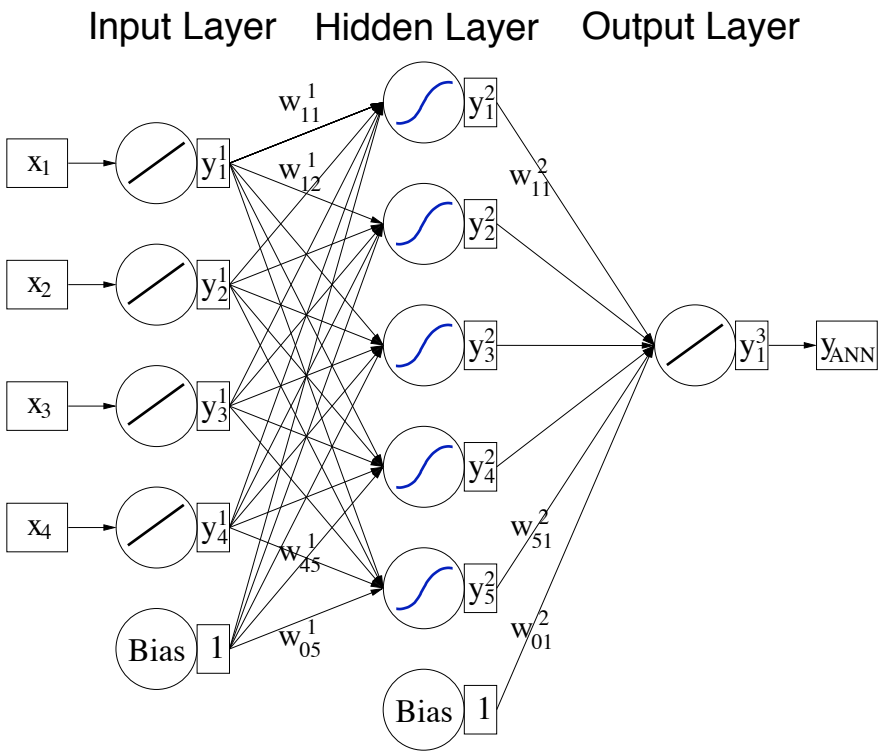

Figure 6.1: Diagram of how neurons may be arranged in an ANN to take a set of four input variables $\{x_1, x_2, x_3, x_4\}$ to a single-valued output, y_{ANN}. Here, an input layer of four neurons, each corresponding to one of the input variables, then passes along a signal to a hidden layer of five neurons, which pass their signals to a one-neuron output layer, which corresponds directly to the ANN output. Figure adapted from [65].

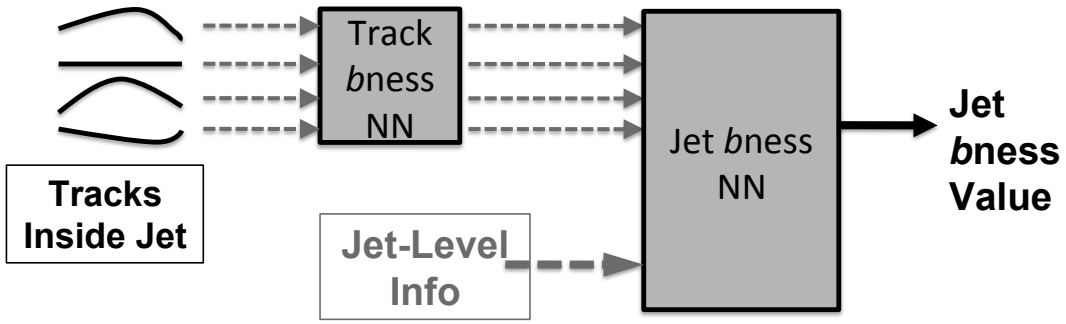

Figure 6.2: A flow chart describing the jet *b*ness NN tagger. We start with individual tracks in a jet, and use various parameters of them as inputs into a track-by-track NN that assigns each track a *b*ness. These values are collected and used, along with other jet-level parameters, as inputs into a jet-by-jet NN. The final output of this NN is called the jet *b*ness.

taggers.

6.2 Description of the *b*ness ANN

The approach behind the development of the jet *b*ness tagger was to focus on the basic observable constituents of *b*-jets: the charged particle tracks located within the jet cone. *b*-jet tagging relies on properties of the *B* hadron decays, and thus a careful look at properties of individual tracks is a strong starting point for identification of *b* jets. Along with looking at individual tracks, we look at properties of groups of tracks (like if tracks come from a secondary vertex) and properties of the jet as a whole.

A flow-chart of how the *b*ness NN works is shown in Fig. 6.2. The *b*ness NN is organized in the following way: each track inside the jet is assigned a "track *b*ness" using a track-by-track NN that assigns a high score to tracks likely coming from *B* hadron decays, and a low score from tracks that likely do not. Information on the tracks is collected and combined with other jet-level variables to be used as inputs into a jet-by-jet NN. The output of this NN is the "jet *b*ness", assigning a high score for jets likely to be coming from

b quarks. We can then place a cut requiring a minimum jet *b*ness value to select events rich in heavy-flavor jets.

6.2.1 Basic track selection

In order to maintain the largest amount of information available to us for *b* tagging, we select tracks to enter the track-by-track NN using a loose criteria, rejecting only those tracks with $p_T < 0.4$ GeV/c (a standard CDF cut), or whose only hits are in the COT, which does not have the impact parameter resolution necessary to find significantly displaced tracks. We only consider tracks located within the cone $R = 0.4$ around the jet axis. Also, we reject tracks likely originating from the decays of other (non-B) long-lived hadrons, like the K_s ($m = 0.497$ GeV/c^2) and the Λ ($m = 1.115$ GeV/c^2). Track pairs that can form a two-track vertex and whose invariant mass lies within 10 MeV of those particles' masses are removed from consideration.

6.2.2 The track neural network

The variables used as inputs to the track-by-track NN are related to two major properties expected in B hadron decays: charged particle tracks displaced from the primary interaction vertex, and tracks with high-p_T located close to the jet axis. The former is due to the long lifetime of B hadrons, while the latter is due both to the typical boost of B hadrons inside the jet and the high mass of B hadrons compared to their decay products. Variables related to the displacement include the following:

- the track's signed impact parameter, d_0, where the sign is positive if the angle between the jet direction and the line joining the primary vertex to the point of closest approach of the track to that vertex is less than 90°, and is negative otherwise;

- the significance of that impact parameter given its uncertainty, d_0/σ_{d_0};
- the track's z displacement from the primary vertex, z_0;
- and, the significance of that z displacement, z_0/σ_{z_0}.

Variables related to track kinematics include:

- the track's transverse momentum, p_T;
- its pseudorapidity with respect to the jet axis, η_{axis};
- and, the magnitude of the track's momentum in a direction perpendicular to the jet axis, p_{perp}.

In addition, the E_T of the jet is included as an input variable in the track-by-track NN, since many of these quantities are strongly correlated with the jet energy. In order to not bias our tagger for jets of a certain energy, B hadron tracks and non-B hadron tracks used for training the NN are weighted so as to have the same parent jet E_T distribution.

These input variables are shown for B hadron tracks and non-B hadron tracks from a PYTHIA $ZZ \to jjjj$ MC sample in Fig. 6.3. Tracks are considered to be from B hadrons if they are matching within $\Delta R < 0.141$ to particles that come from the decay a B hadron in the simulation's event record. As expected, tracks from B hadrons show higher displacement, higher p_T, and are closer to the jet center. The displacement variables—particularly the d_0 and d_0/σ_{d_0}—give the highest level of discrimination.

The track-by-track NN is a multi-layer perceptron (MLP) trained using the TMVA [65] package. The NN is trained using tracks from jets from $ZZ \to jjjj$ decays in MC, and utilizes two hidden layers of 15 and 14 nodes. The single-valued output is called the "track *b*ness", which ranges from -1 to 1, and examples of which are shown in Fig. 6.4.

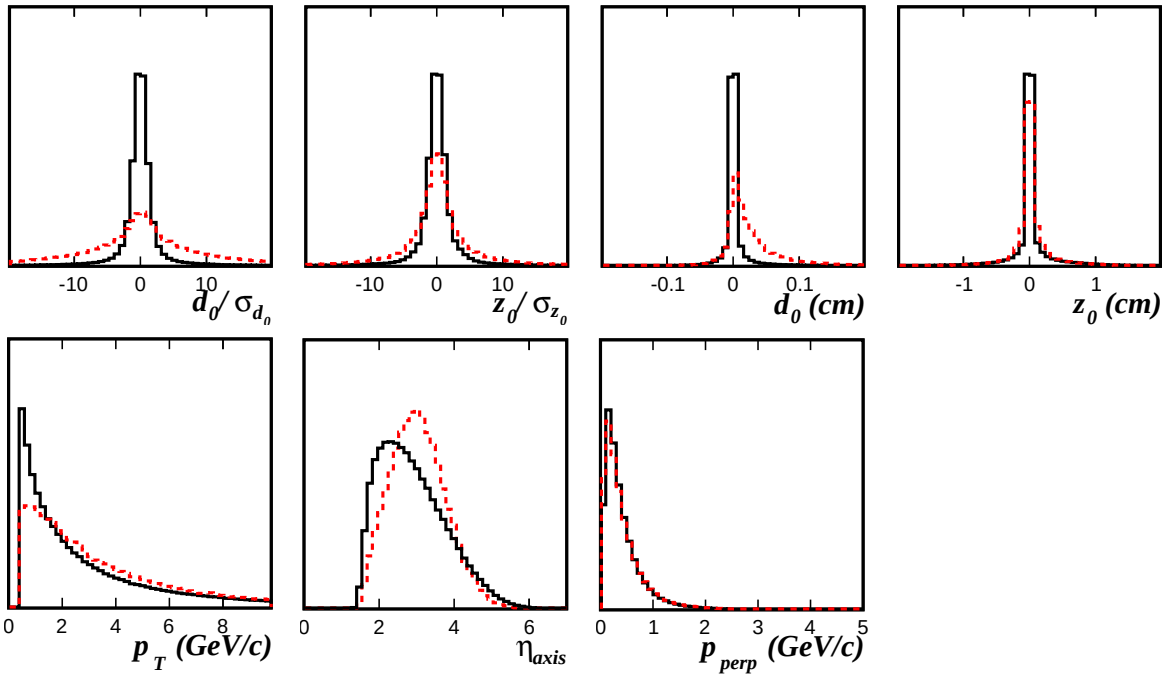

Figure 6.3: Inputs used in the neural network for calculating the per-track *b*ness for tracks from *B* hadron decays (*dashed red lines*), and tracks not from *B* hadron decays (*solid black lines*). The *y*-axis in each plot is in arbitrary units, and the distributions from *B* hadron daughter tracks and non-*B* hadron daughter tracks are normalized to the same area.

6.2.3 The jet neural network

In identifying jets most likely to come from b quarks, the most important properties we consider are those related to the tracking information. b jets contain tracks from B hadron decays, so they should have a higher number of high-$b$ness tracks, and those tracks should be consistent with originating from a long-lived B hadron decay. We consider the following variables related to the tracks:

- the track $b$ness values of the five most b-like tracks in the jet, b_i, $i = 0 \ldots 4$;
- the number of tracks with track $b$ness > 0, n_{trk};
- the significance of the displacement of a reconstructed secondary vertex from the primary vertex, $L_{xy}/\sigma_{L_{xy}}$, where the secondary vertex is fit using tracks with track $b$ness > -0.5;
- and, the invariant mass of the tracks used to fit the secondary vertex, m_{vtx}.

While less powerful for identifying b jets, additional factors are also considered:

- the number of K_s meson candidates found inside the jet, as b jets tend to contain more K_s mesons than non-b jets;
- and, if a candidate for a soft muon is found inside the jet, the likelihood that object is truly a muon calculated by a soft muon tagger [64], as some b jets contain semi-leptonic decays.

The above variables, along with the jet E_T, are used as inputs into the jet-by-jet NN. It is trained in a similar fashion to the track-by-track NN, with two hidden layers of 15 and 16 nodes. The network is trained on jets from a $ZZ \to jjjj$ MC sample, separating jets matched to b quarks against all other types of jets. The most important variables in the jet-by-jet NN, those related to the track properties, are shown in Fig. 6.4.

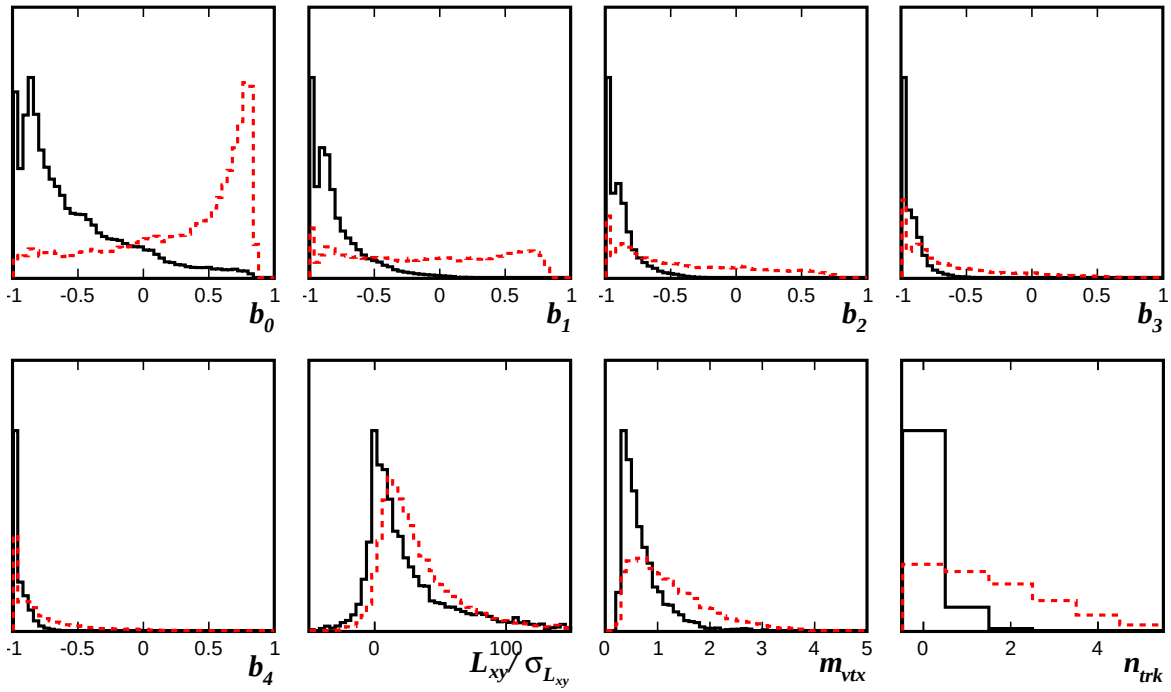

Figure 6.4: Inputs used in the neural network for calculating the per-jet *b*ness for *b* jets (*dashed red lines*), and non-*b* jets (*solid black lines*). The *y*-axis in each plot is in arbitrary units, and the distributions from *b* jets and non-*b* jets are normalized to the same area.

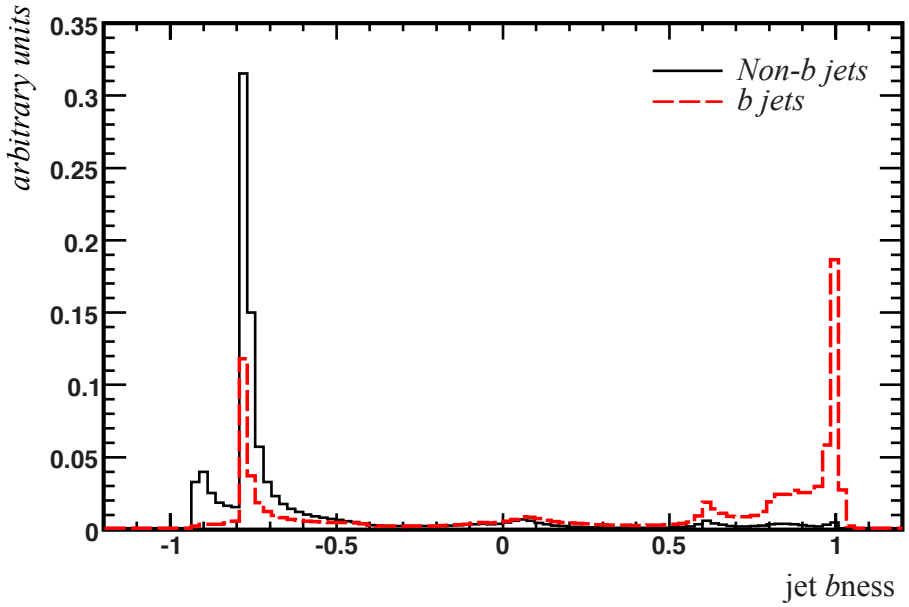

Figure 6.5: Output of the final jet-by-jet NN, the jet *b*ness, for *b* jets (*dashed red line*) and non-*b* jets (*solid black line*).

The final output of the jet-by-jet NN, the "jet *b*ness", is shown in Fig. 6.5 for *b* jets and non-*b* jets from the *ZZ* sample used for training. The choppiness of the distribution is due to the use of discrete-valued variables in the jet-by-jet NN. A number of *b* jets have low jet *b*ness values, generally due to a lack of high *b*ness tracks in these jets either because the *B* hadron decayed relatively quickly, or the jet was at higher η where fewer tracks are reconstructed. Particularly, the region near a jet *b*ness of -0.8 is dominated by jets with zero tracks having positive track *b*ness, zero K_S candidates found, and no secondary vertex. The significant number of *b* jets at these low *b*ness values indicates some *b* jets are nearly indistinguishable from non-*b* jets. However, we see a high purity at high jet *b*ness, and we can use a *b*ness threshold requirement to get a heavy-flavor enhanced jet selection that sacrifices some *b* jet acceptance.

6.3 Evaluation of Mistag Rate and Efficiency

In order to use this new b tagger in analyses, we determine the true efficiency of identifying a b jet and the false tag ("mistag") rate for tagging a non-b jet as a function of a minimal $b$ness requirement, $e(b)$ and $m(b)$ respectively. Also, we evaluate the efficiency and mistag rate in Monte Carlo simulations ($e_{\mathrm{MC}}(b)$ and $m_{\mathrm{MC}}(b)$, respectively). By checking differences between these values in data and MC, we can correct our simulation to better match the data.

For any given selection of data, we can calculate the mistag rate (where all non-b jets are considered mistags) if we know the number N_B of b jets, the number $N_B(b)$ of b jets above the threshold $b$ness, the total number N of jets, and the total number $N(b)$ of jets above the $b$ness cut threshold:

$$m(b) = \frac{N(b) - N_B(b)}{N - N_B} \ . \tag{6.1}$$

We may use MC to determine the fraction f_B of all jets that are b jets, and the efficiency $e_{\mathrm{MC}}(b)$ for these jets to pass the $b$ness cut. This efficiency will need to be modified by a scale factor $s_e(b) = e(b)/e_{\mathrm{MC}}(b)$ if it is different from the true efficiency evaluated in data. Thus,

$$N_B = f_B N \ \text{ and } \ N_B(b) = s_e(b) e_{\mathrm{MC}}(b) f_B N \ . \tag{6.2}$$

Also, if we define a mistag rate that has not been corrected for the possible presence of b jets in the same sample, $m_{\mathrm{raw}}(b) = N(b)/N$, then we may write equation 6.1 in the following way:

$$\begin{aligned} m(b) &= \frac{m_{\mathrm{raw}}(b) N - s_e(b) e_{\mathrm{MC}}(b) f_B N}{N - f_B N} \\ &= \frac{m_{\mathrm{raw}}(b) - s_e(b) e_{\mathrm{MC}}(b) f_B}{1 - f_B} \ . \end{aligned} \tag{6.3}$$

We can write an analogous expression for the efficiency of b jets passing a given $b$ness cut,

$$e(b) = \frac{e_{\text{raw}}(b) - s_m(b) m_{\text{MC}}(b) f_L}{1 - f_L} , \tag{6.4}$$

where $e_{\text{raw}}(b)$ is a "raw" efficiency uncorrected for the presence of non-b jets in a sample, $m_{\text{MC}}(b)$ is the mistag rate as measured in MC, corrected to match data by a scale factor $s_m(b)$, and f_L is the fraction of light-flavor (here just simply defined as non-b) jets in the chosen sample. Both $e_{\text{raw}}(b)$ and $s_m(b) m_{\text{MC}}(b) = m(b)$ can be calculated easily by counting events above a given $b$ness threshold in the data and MC respectively.

We will use a $t\bar{t}$ selection to determine the efficiency. Because of the different competing processes in this $t\bar{t}$ sample (there is a significant contribution from W + light flavor jets and $W + b\bar{b}$ processes), it is best to break f_L into these most significant subsamples,

$$f_L = \frac{f_L^{Wjj} N_{Wjj} + f_L^{Wb\bar{b}} N_{Wb\bar{b}} + f_L^{t\bar{t}} N_{t\bar{t}}}{N_{Wjj} + N_{Wb\bar{b}} + N_{t\bar{t}}} , \tag{6.5}$$

where N_X is the number of events predicted by MC in subsample X, and f_L^X is the fraction of non-b jets in subsample X. We assume that the MC correctly reproduces the values of f_L^X. To determine $s_e(b) = e(b)/e_{\text{MC}}(b)$, we write down a similar expression for the efficiency in MC using the efficiency of each subsample in MC,

$$e_{\text{MC}}(b) = \frac{1}{N_{Wjj} + N_{Wb\bar{b}} + N_{t\bar{t}}} \sum_X e_X(b) f_B^X N_X , \tag{6.6}$$

where, as before, N_X is the number of events predicted by Monte Carlo in subsample X, f_B^X is the total fraction of b jets in subsample X, and e_X is the efficiency of b jets passing a particular $b$ness cut in subsample X. We assume, again, that the Monte Carlo correctly reproduces the values of f_B^X.

By calculating the mistag rates and efficiencies in this way, we have removed any di-

rect dependence on the number of events. This removes uncertainties that purely effect the normalization, like the uncertainty on the total integrated luminosity. Also, the determination of the mistag rate depends on the calculated value of the efficiency (through the scale factor term $s_e(b)$), and that in turn the determination of the efficiency depends on the mistag rate (again through the scale factor $s_m(b)$). Similarly, the uncertainties on these quantities (see below) depend on each other. Thus, we use an iterative procedure to solve for the mistag rate, efficiency, and their uncertainties. We calculate the mistag rate first using a value of $s_e(b) = 1$, and find that the values of $e(b)$ and $m(b)$ and their uncertainties converge very quickly.

The uncertainties on these quantities may also be calculated from the expressions above. For the mistag rate,

$$\begin{aligned}\sigma_m^2(b) = &\frac{m_{\text{raw}}(b)(1 - m_{\text{raw}}(b))}{N(1 - f_B)^2} \\ &+ \left(\frac{\sigma_e(b) f_B}{1 - f_B}\right)^2 \\ &+ \left(\frac{\sigma_{f_B}[s_e(b)e(b) - m(b)]}{1 - f_B}\right)^2 . \end{aligned} \tag{6.7}$$

The first term is a binomial uncertainty on the raw mistag rate of the sample, and is the term related to the statistical uncertainty of the sample used to determine the mistag rate. The second term comes from the uncertainty on the measured value of $e(b)$ (see Eq. 6.8). The final term is due to the uncertainty on f_B, which will depend on the choice of MC and the region in which MC and data are compared.

Given the multiple processes we must consider in our $t\bar{t}$ selection (see Eqs. 6.5 and 6.6), the expression for the uncertainty on the efficiency is more complicated. To ease matters, we obtain this uncertainty by calculating the uncertainty of the quantity ($e(b) -$

$e_{\mathrm{MC}}(b)$), and find

$$\sigma_e^2(b) = \frac{1}{(1-f_L)^2}\left(\frac{e_{\mathrm{raw}}(1-e_{\mathrm{raw}})}{N_D} + (\sigma_m f_L)^2\right) + \sum_X \frac{\sigma_X^2}{[N_{\mathrm{MC}}(1-f_L)]^2} \times \left[(e+s_m m)(f_L - f_L^X) + f_B^X(e_{\mathrm{MC}} - e_X)\right]^2 , \quad (6.8)$$

where the latter term represents a sum over each of the MC subsamples. N_{MC} and N_B are the total number of events and events with b jets in the MC, and σ_X is the uncertainty assigned to the number of events in each MC subsample. Because we compare only the normalizations of data and MC in our determination of efficiency (and mistag rate) scale factors, the uncertainty on the number of events in each MC subsample need only reflect the relative uncertainty on the fraction of events each subsample contributes to the whole. We assign $\sigma_{Wb\bar{b}}$ = 20%, and σ_{Wjj} = 8.72% and $\sigma_{t\bar{t}}$ = 6.78% based on a fit to the distribution of the sum of the highest two $b$ness jets in $t\bar{t}$ events.

6.4 Selection for Mistag Rate and Efficiency Determination

Following the procedure described in Sec. 6.3, we must choose two independent regions in which to determine the mistag rate and efficiency of the b tagger. To reduce uncertainties, it is best to choose a well-modeled region dominated by falsely tagged jets (where we expect few b jets) and a well-modelled region rich in b jets. For the former, we choose events containing two oppositely charged electrons or muons likely from the decay of a Z boson, plus one jet. The Z + 1 jet selection closely follows that described in Chap. 4, and are summarized in Tab. 6.1. For the latter, we choose events containing the decay of a pair of top quarks, where we require exactly one lepton, at least four jets, and a large imbalance in transverse momentum in the event, indicating the likely presence of a neu-

trino. We expect that the two jets with the highest *b*ness values in this sample will very likely be *b* jets. The cuts applied for the $t\bar{t}$ selection region is described in Tab. 6.2.

Z + 1 *jet Selection*
N_{leptons} = 2, both electrons or both muons
Leptons have opposite charge
Δz_0 between leptons < 5 cm
Lepton $p_T > 20$ GeV/c
75 GeV/c^2 $< M_{ll} <$ 105 GeV/c^2
$\not\!\!E_T < 25$ GeV
Reconstructed $p_T(Z) > 10$ GeV/c
$N_{\text{jets}}(E_T > 10 \text{ GeV}) = 1$
Jet $E_T > 20$ GeV, $\lvert\eta\rvert < 2.0$

Table 6.1: Summary of event selection requirements for the Z + 1 jet sample used to determine the mistag rate of the jet *b*ness tagger.

$t\bar{t}$ *Selection*
N_{leptons} = 1
Lepton $p_T > 20$ GeV/c
$\not\!\!E_T > 20$ GeV
$\not\!\!E_T$-significance $> 1(3)$ for $\mu(e)$ events
Reconstructed $M_T(W) > 28$ GeV/c^2
Highest two *b*ness jets' $E_T > 20$ GeV
$N_{\text{jets}}(E_T > 15 \text{ GeV}) \geq 4$
Total sum $E_T > 300$ GeV

Table 6.2: Summary of event selection requirements for the $t\bar{t}$ sample used to determine the tag efficiency of the jet *b*ness tagger. The total sum E_T is defined as the sum of the lepton p_T, $\not\!\!E_T$, and E_T of all jets with $E_T > 15$ GeV.

These events are selected by specific high-p_T central electron and muon triggers that do not have additional jet selection requirements. We use data corresponding to an integrated luminosity of 4.8 fb^{-1}. We use ALPGEN interfaced with PYTHIA for parton showering, to model W and Z plus jets samples and PYTHIA to model $t\bar{t}$ and other processes with small contributions. Similar to how we determine the trigger and lepton ID/reconstruction

efficiencies for our $Z + 2$ jet sample by comparing data and MC in a $Z + 1$ jet sample, we check the trigger and ID/reconstruction efficiencies against a sample of $Z \to e^+e^-$ or $\mu^+\mu^-$ events without jets. Table 6.3 contains a summary of the total number of events.

	Electrons	Muons
$Z + 1$ jet selection		
Data Events	9512	5575
MC Events	9640 ± 880	5540 ± 490
$t\bar{t}$ Selection		
Data Events	507	835
MC Events	542 ± 56	862 ± 85

Table 6.3: Number of events in data and MC in the $Z + 1$ jet selection region, after proper scale factors have been applied. The uncertainties on the MC reflect only the two dominant systematic uncertainties: the uncertainty on the jet energy scale and the uncertainty on the luminosity. Overall, the agreement in number of events is good.

6.5 Mistag Rate Determination

Fig. 6.6 shows the jet *b*ness distribution for jets in the $Z + 1$ jet sample. The sample is dominated by light-flavor jets, but there is a significant contribution of real *b* jets at higher *b*ness values, coming from $Z + b\bar{b}$ production. The *b*-jet incidence rate reaches above 60% for the highest *b*ness cuts, and thus we will expect the uncertainties in the mistag rate to be substantially higher there, due to both the small sample of available jets and the high contamination rate combined with the uncertainty on the number of *b* jets in that smaller sample.

Using the method described in Sec. 6.3, we determine the values of $m(b)$ and the relative difference between the mistag rate in data and MC ($s_m(b) - 1$). We also calculate the uncertainty on the mistag rate. The value of the scale factor and its uncertainty at the relevant *b*ness cut in the ZW/ZZ search is summarized in Tab. 6.4.

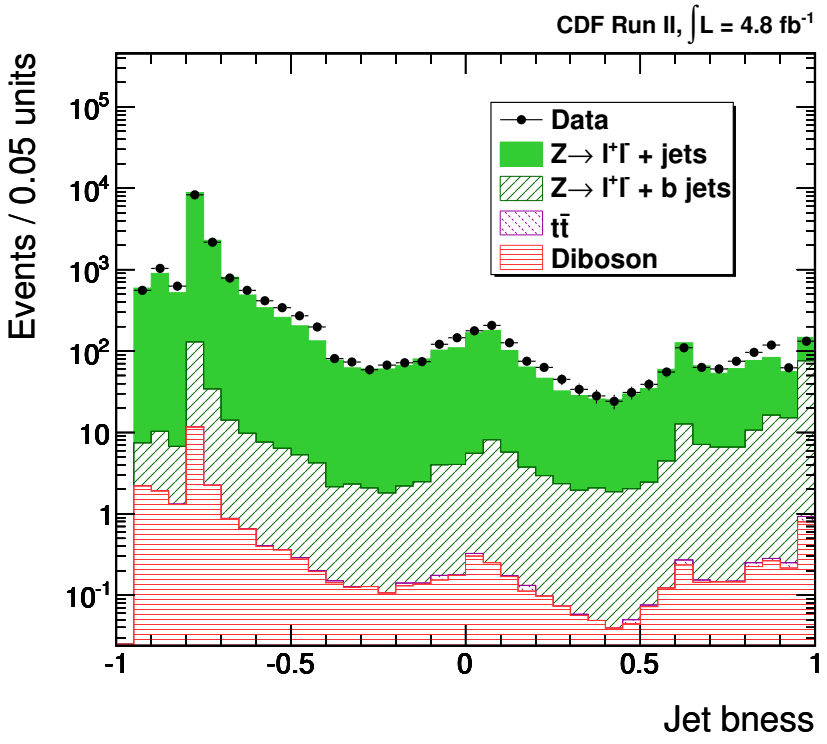

Figure 6.6: A comparison of the jet *b*ness in data and MC in the $Z+1$ jet selection region. The MC is able to reproduce the main features of the *b*ness distribution in data. We use this distribution to determine the mistag rate for placing a cut on jet *b*ness in data, and use the differences between data and MC to determine corrections to the mistag rate in MC.

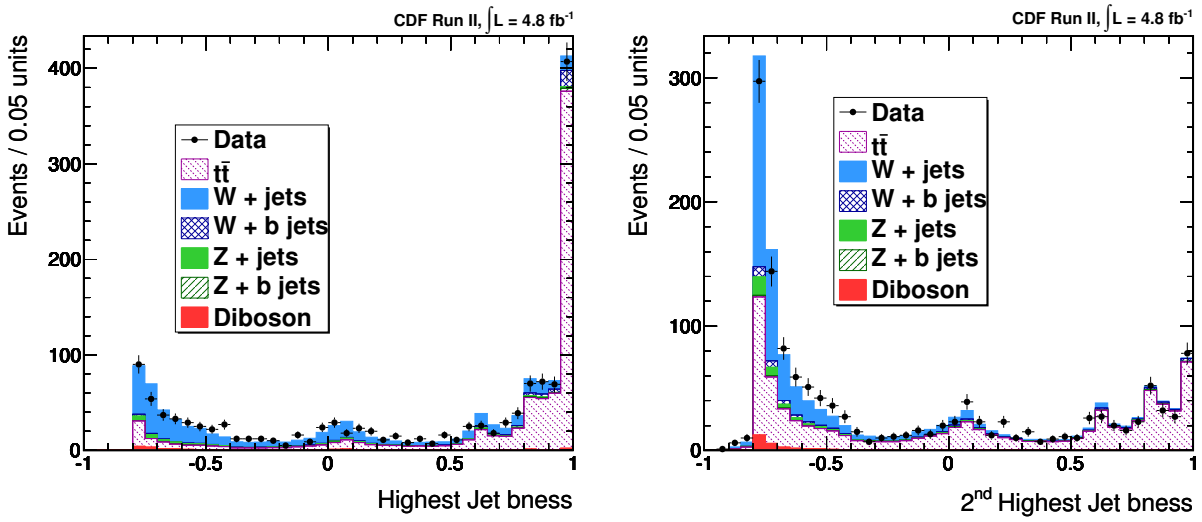

Figure 6.7: Jet *b*ness of the first (left) and second (right) jet, as ordered by *b*ness, in the $t\bar{t}$ lepton + jets selection region. The simulation reproduces most of the features of the data, and we see much of the *b*-enriched samples clustered towards high *b*ness.

6.6 Tagging Efficiency Determination

We use our $t\bar{t}$ selection, described in Sec. 6.4 and Tab. 6.2, to calculate the efficiency from a sample of jets with high *b* purity. As these events have many jets, we order the jets by decreasing *b*ness value. Fig. 6.7 shows the jet *b*ness distributions in data and MC for the two jets with highest *b*ness in each event. The agreement here is very good, and regions of high *b*ness are almost exclusively populated by $t\bar{t}$ events, indicating that our *b* tagger is properly identifying *b* jets.

We calculate the efficiency of a given *b*ness threshold and its uncertainty as described in Sec. 6.3. The relative differences and uncertainties on the efficiency are on the order of 10% or less (see [61]). Tab. 6.4 lists the efficiency and mistag rates in data and MC for a chosen operating point—the 2$^{\text{nd}}$ highest jet *b*ness > 0.0—along with the relative difference between data and MC, and the error on that difference. We find that the MC simulation typically underestimates the mistag rate while overestimating the tagging effi-

ciency. More direct comparisons of the performance of the jet *b*ness tagger with CDF's SecVtx tagger are included in [61].

Quantity	*b*ness Cut	Data	MC	% Difference	% Error
Mistag Rate	> 0.0	0.0819	0.0720	14%	4.1%
Tag Efficiency	> 0.0	0.622	0.684	-9.0%	8.7%

Table 6.4: Mistag rates and efficiencies on a 2^{nd}-highest jet *b*ness cut > 0.0 determined from comparisons of data and MC in our $Z + 1$ jet and $t\bar{t}$ control regions.

While generic comparisons between taggers are difficult, we compare our tagger to the most commonly used *b* tagger in CDF, the SecVtx tagger [62]. The efficiency and mistag rates of our tagger compare favorably to the SecVtx tagger. We compare the two taggers using simulated events, looking at the two highest *b*ness jets in the MC of our $t\bar{t}$ selection. The "tight" SecVtx tagger operating point on this sample of jets has an efficiency of 0.59 and a mistag rate of 0.052, while the "loose" operating point has an efficiency of 0.68 with a mistag rate of 0.088. For a 2^{nd} highest *b*ness jet cut at > 0.0, we have a similar efficiency to the "loose" SecVtx tag with a slightly lower mistag rate.

CHAPTER 7

NEURAL-NETWORK-BASED QUARK/GLUON DISCRIMINANT

In this analysis, we search for two high-p_T leptons from the decay of a Z boson, and two jets from a $W \to q\bar{q}'$ or $Z \to q\bar{q}$ decay; thus, the two jets in our signal are quark jets. Our dominant background, two jets produced in association with a $Z \to \ell^+\ell^-$ decay, contain a significant number of gluon jets. Thus, similar to the tagging of b jets, our sensitivity to $W^\pm \to q\bar{q}'$ and $Z \to q\bar{q}$ decays may be increased if we can tag jets likely to come from light quarks—u, d, s, and the slightly heavier c—and separate them from jets likely originating from the fragmentation of gluons.

Gluon jets, due to their higher color charge, tend to contain a higher particle multiplicity and be spatially broader in the detector than quark jets [58]. We attempt to quantize the spatial spread of jets using a collection of artificial neural-networks, trained to separate gluon jets from light-flavor quark jets. We call the result of the final ANN the jet quark/gluon value (or jet QG value). We may place a cut on this value to use it as a light-flavor quark tagger. Our approach is unique in that we look directly at bulk distributions of the individual towers and tracks within a jet by constructing the distribution of distances between pairs of these objects.

We calibrate the agreement of response of the final ANN in MC to the response in data using a $W \to \ell\nu$ + 1 jet sample. Similar to our characterization of the jet $b$ness variable as a tagger, we determine a tagging efficiency and mistag rate of placing a cut on the jet QG value using two independent samples: $W \to \ell\nu$ + 2 jets, similar to our Z + jets background, and $t\bar{t} \to b\bar{b}\ell\nu q\bar{q}'$, which contains two jets from the hadronic decay of a W boson, very much like our diboson signal.

7.1 Introduction

The strong force, and the behavior of the quarks and gluons that compose its interactions, is described by Quantum Chromodynamics (QCD). An essential property of QCD is that while quarks possess color, that only color-neutral states exist in nature. The transformations of the color states of quarks are defined by group SU(3), with the eight unitary matrices that generate that group giving rise to the eight gluons that mediate the strong force. These gluons have an effective color charge that is larger than that of the quarks. The color charge is proportional the square root of the ratio of the "color factors": C_A and C_F, repsectively. In SU(3), $C_A = 3$ and $C_F = 4/3$, thus $C_A/C_F = 9/4$.

This increased color charge for gluons results in a number of properties that distinguish the hadronization of quark and gluon jets [58]. First, the multiplicity of any type of final-state object should be higher in gluon jets: the largest contributor to increases in multiplicity is the radiation of gluon jets, and since the effective gluon color charge is higher, gluons are likely to radiate more. As a result of this higher multiplicity, the final-state particles in gluon jets should have a softer momentum spectrum than quark jets of a similar energy. In addition to increases in particle multiplicities in gluon jets, gluon jets will also tend to be broader, spatially, than quark jets. The jet width tends to increase with decreasing energy, and since the energies of partons undergoing hadronization in gluon jets are typically lower, gluon jets will tend to have larger angular size than quark jets of a similar energy. These properties, however, are not suitable for separating b quarks from gluon jets, as the weak decays of B hadrons will lead to different properties.

Experimental studies of the differences between gluon and quark jets are difficult, as it is hard to unambiguously determine whether a jet originated from a quark or gluon. AMY, a detector located at an e^+e^- storage ring at the KEK laboratory in Japan, collected events with two back-to-back jets, likely from quarks, and events with three jets, which

likely includes a gluon radiated from a quark jet [69]. They determined that gluon jets have less energy inside a core region of the jet, as compared to light quark jets, indicating a wider angular spread for gluon jets. OPAL, an experiment located at the e^+e^- LEP (Large Electron-Positorn) collider at CERN, was able to obtain a high purity of gluon jets from three-jet decays of the Z boson: $Z \to b\bar{b}g$ [70]. The leading-E_T jet is very likely a b jet, and by b-tagging one of the two other jets, one may take the untagged jet to be from a gluon. OPAL found that gluon jets had softer spectrum of particle energies than light quark jets, as predicted by the theory.

Consideration of jet widths, to reduce backgrounds from gluon jets to final states with only quark jets, has been used in analyses at CDF. Searches for all-hadronic decays of top quark pairs (where both W bosons decay to pairs of quarks) [71] and all-hadronic decays of WH/ZH production [72] included evaluations of the η and ϕ moments of jets, each a single-value measurement of the width of the jet, in ANN-based discriminants to reduce contributions from multijet backgrounds. However, these distributions have not been used as simple tagging variables, upon which one may place a cut, because they have little discrimination power on their own. Our approach is to develop a variable that is specifically formulated to distinguish between quark and gluon jets, similar to the b-taggers described in Chap. 6.

7.2 Jet QG Value Definition

There are two basic detector-level objects associated with jets at CDF: calorimeter towers, which are clustered together to form the actual jets, as described in Sec. 3.3; and, charged particle tracks, which may be associated with a jet via matching in ΔR, even if they are not used to define the jets themselves. These two sets of objects are in many ways complementary to describing properties of the jet. Charged particle tracks are reconstructed

with high spatial resolution, and, if including information from the silicon detectors, provide the "earliest" description of development of the jet; however, the tracking systems provide limited coverage for $|\eta| > 1$, and even inside that region only relate information about a fraction of the particles in the jet—those that have charge. Calorimeter towers provide a more complete picture of the jet, giving energy measurements of both charged and neutral particles over a wider η range; but, we cannot determine particle multiplicities from the calorimeter towers, and the large tower size (typically 0.1 in $\eta \times 15^\circ$ in ϕ) gives us less resolution on measurements of the jet width.

For this reason, we use information from both tracks and towers to form our quark/gluon discriminant. The discriminant itself uses a total of three ANNs (see Sec. 6.1 for a description of how our ANNs work), described visually in a flow-chart in Fig. 7.1. There are two networks for separating quark and gluon jets by looking at the distribution of energy contained in calorimeter towers and the distribution of momenta contained in reconstructed charged-particle tracks, described in the following sections. Most jets may be assigned a Tower NN value and Track NN value, the output of these two ANNs. These two NN values are combined in a third ANN, along with other variables that offer some discrimination between quark and gluon jets, or that are related to the how spread or collimated jet it is. The output of this final ANN is the jet QG value, and is our final discriminant for separating quark and gluon jets.

Each of the ANNs is trained on jets matched to a light flavor (*uds*) quark or gluon with $p_T > 20$ GeV/c and within $\Delta R = 0.4$ of the center of the jet, with no other partons above 8 GeV/c within $\Delta R = 0.7$. The jets come from a $Z \rightarrow \mu^+\mu^-$ + 2 parton ALPGEN sample, interfaced with PYTHIA showering. Similar to the jet *b*ness ANNs, each ANN in the quark/gluon discriminant is a feed-forward multilayer perceptron with a single output, implemented using the MLP algorithm from the TMVA package [65]. The networks are trained on 100,000 quark and gluon jets, and tested for biases in over-training on a sam-

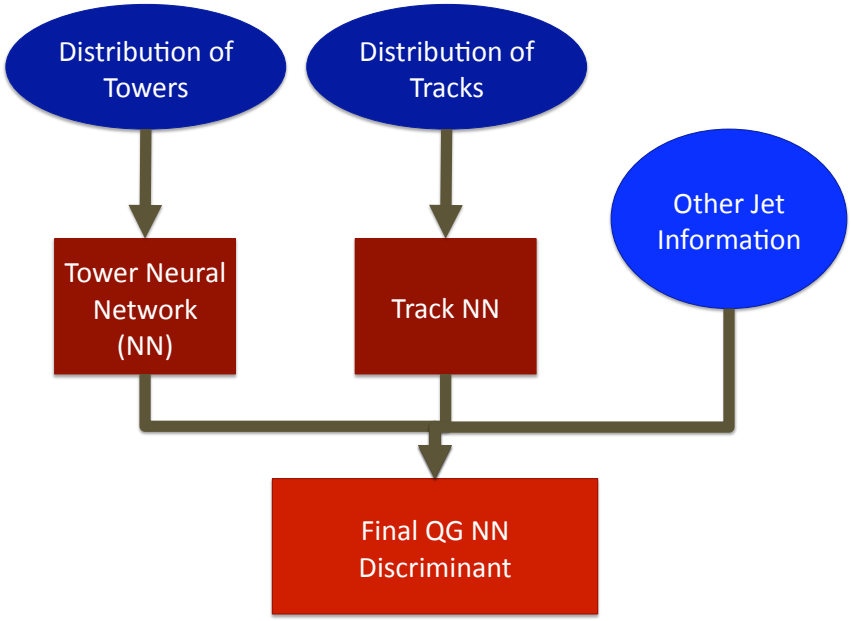

Figure 7.1: A flow chart describing the jet QG value quark/gluon discriminant. We start by forming a distribution of the towers and tracks inside a cone of $R = 0.7$ around each jet. These distributions are used as inputs into two separate ANNs, giving us a "Tower NN" value and a "Track NN" value. These are then used as inputs, along with other variables describing aspects of the jet, into a final NN. The output of this final network is called the "jet QG value".

ple with 500,000 quark and gluon jets. Gluon jets are reweighted to match the E_T and η distributions of the quark jets to remove any bias these variables.

7.2.1 The tower neural network

For each jet, we obtain a list of the calorimeter towers (with $E_T > 1$ GeV) in a cone of $R = 0.7$ around that jet. Each tower has a location coordinate, (η, ϕ), and deposited energy E associated with it. We construct a distribution of the distance, $\Delta R_{ij} = \sqrt{(\Delta\eta_{ij})^2 + (\Delta\phi_{ij})^2}$, between all pairs of towers within the jet. We weight each tower pair by its relevance in terms of energy, resulting in a distribution that characterizes the spatial spread of the energy within each jet. The weight we apply to each tower pair is given by

$$\frac{E_i E_j}{0.5((\Sigma E)^2 - \Sigma E^2)} \,,$$

where E_i and E_j are the energies of the two towers in the pair, ΣE is the sum of the energy in all towers in a cone of $R = 0.7$ around the jet, and ΣE^2 is the sum of the square of the energy of each tower in that same cone. This denominator is chosen in order to normalize the sum of all weights of tower pairs to unity. In this thesis, this distribution is also referred to as the ΔR_{towers} distribution.

The typical ΔR_{towers} distribution for quark and gluon jets is shown in Fig. 7.2. The x-axis is the distance between a pairs of towers around the jet, while the y-axis is the typical energy content contained in tower pairs at that distance. We see that light quark jets have a higher proportion of their energy clustered together, with the ΔR_{towers} peaking lower than for gluon jets.

We split the ΔR_{towers} distribution into 56 bins with bin size $\Delta R_{\text{towers}} = 0.025$ for $0.0 \leq \Delta R_{\text{towers}} \leq 1.4$, and the contents of the 53 non-zero bins (the first 3 bins are empty due to the segmentation of the calorimeter) are used as inputs for the Tower NN. This

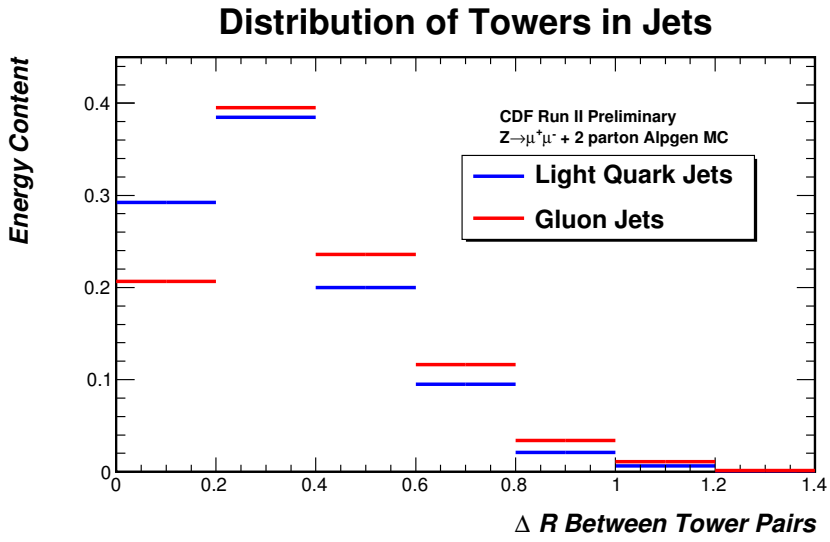

Figure 7.2: A typical distributions of ΔR between pairs of towers in light-flavor quark (blue) and gluon (red) jets. Light-flavor quark jets tend to peak at low ΔR_{towers}, indicating they are rather collimated, while gluon jets tend to have a higher ΔR_{towers} distribution.

bin size was chosen to correspond to the maximum resolution for $\Delta R_{\text{towers}} = 2.0$, as we originally included information out to a cone-size of $R = 1.0$, but found that it offered little improvement over a cone size of $R = 0.7$. However, extending the information included beyond the jet size of $R = 0.4$ to $R = 0.7$ does provide improved discrimination power because gluon jets are more likely to radiate some energy outside the smaller cone size.

In constructing the quark/gluon discriminant, we also considered including a second ΔR distribution, one populated by the distance of each individual tower from the jet center. However, the information from this distribution is largely contained in the ΔR_{towers} distribution, and so it offers little additional discrimination at a price of added complexity of the network. Also, a benefit of including the distance between tower pairs over the distance between each tower and the jet center is the increased multiplicity of tower pairs: if there are n towers included in a jet, then there are $n \cdot (n-1)$ possible tower pairs, versus n tower-jet center pairings. This greater multiplicity means the ΔR_{towers} distribution will

be more populated, providing greater discrimination power. For these reasons only the contents of the ΔR_{towers} distribution are used as inputs into the Tower NN.

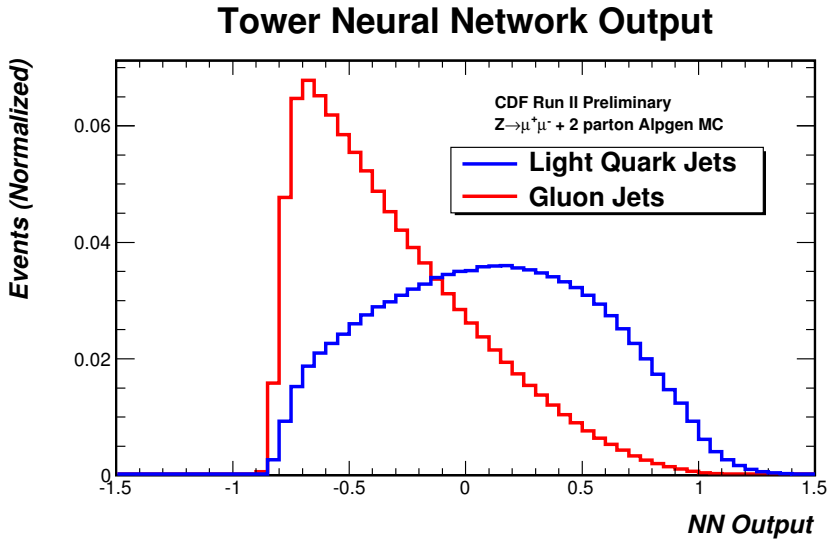

Figure 7.3: The output of the Tower ANN discriminant for light-flavor quark (blue) and gluon (red) jets. We see a good level of discrimination between quark and gluon jets in this NN variable.

The output of the Tower NN for quark and gluon jets from our $Z \rightarrow \mu^+\mu^-$+ jets MC is shown in Fig. 7.3. We see that, in the MC, the distribution of gluon jets peaks at low values, while the distribution of Tower NN values is more rounded and peaks at a higher value. We see a smoother behavior here than in the output of the *b*ness NN (see Fig. 6.5) due to the lack of integer-valued inputs. While the NN is trained to produce an output between -1 and +1, some jets appear at Tower NN values > 1 due to differences from these jets and the training sample. However, the distributions still look well-behaved at these values. Overall, we see good discrimination between quark and gluon jets in this Tower NN value.

7.2.2 The track neural network

When looking at the tracks inside a jet, we follow a similar prescription. We collect tracks within a cone of R = 0.7 around each jet, using the tracks' locations in (η, ϕ) (at the primary vertex) and momenta p to obtain a distribution of the distance between pairs of tracks (in ΔR), with each pair weighted by the momentum contained within that pair,

$$\frac{p_i p_j}{0.5((\Sigma|p|)^2 - \Sigma|p|^2)} \text{ ,}$$

with similar definitions for $\Sigma|p|$ and $\Sigma|p|^2$ as in tower energies. We require all tracks come within 5 cm of the reconstructed primary vertex, and that the tracks otherwise meet the requirements outlined in Sec. 6.2.1.

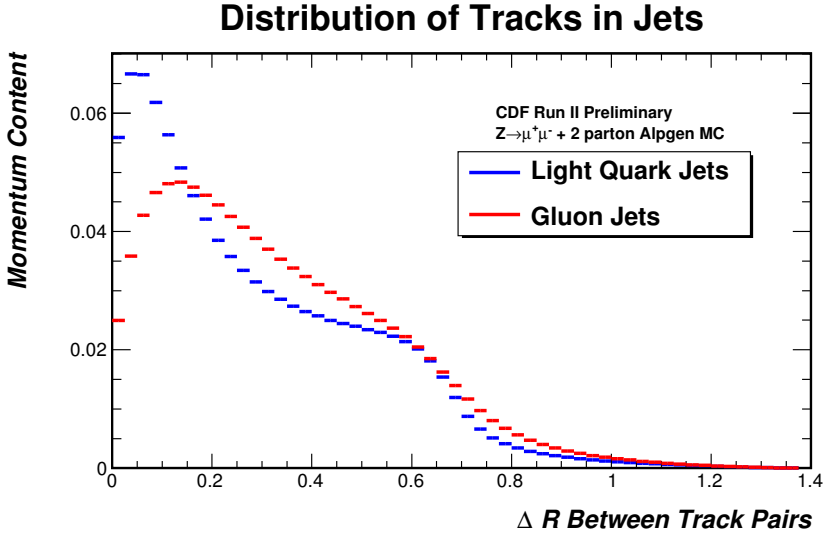

Figure 7.4: A typical distributions of ΔR between pairs of tracks in light-flavor quark (blue) and gluon (red) jets. Similar to the ΔR_{towers} distribution, light-flavor quark jets tend to peak at low ΔR_{tracks} while gluon jets tend to have a higher ΔR_{tracks} distribution.

The typical ΔR_{tracks} distributions for quark and gluon jets (with two or more tracks) are shown in Fig. 7.4. We show more bins in this distribution than in Fig. 7.2, as the

segmentation of the calorimeter makes showing more bins visually difficult to read. The behavior in ΔR_{tracks} is similar to that in the ΔR_{towers} distribution: light quarks jets peak at lower ΔR_{tracks} than gluon jets, indicating more of the track momentum content is in track pairs close together in quark jets compared to gluon jets. The ΔR_{tracks} distribution does not fall with constant rate, with what looks like a second peak appearing at $\Delta R \approx 0.6$, due to geometric effects based on the cone size of the jet.

We split the ΔR distribution between track pairs (or ΔR_{tracks}) into the same 56 bins as used in the Tower NN, and the content of each bin is used as an input into the Track NN. This NN is independent of the Tower NN; while it is possible to use both the bins of ΔR_{tracks} and ΔR_{towers} in the same NN, the possible correlations between similar bins in tower and tracks does not provide enough additional discrimination power to inspire the complications of training a 109 variable NN. Instead, the Track NN is independent of the Tower NN.

The output of the Track NN for quark and gluon jets with two or more tracks from our $Z \rightarrow \mu^+\mu^-$+ jets MC is shown in Fig. 7.5. Like the Tower NN output, we see that the track distribution has good discrimination power for identifying quark jets from gluon jets. Compared to the Tower NN output, the Track NN output peaks at higher NN score for quark jets, but has a somewhat broader distribution in gluon jets. Unlike the Tower NN distribution, we do see some discontinuous features in the Track NN output. These arise from jets that have only two tracks that meet our quality requirements in them, and so only one bin of the ΔR_{tracks} distribution is non-zero. These jets are more likely to originate from quarks than from gluons, due to the higher particle multiplicity in gluons, hence the presence of these discontinuities at higher Track NN values.

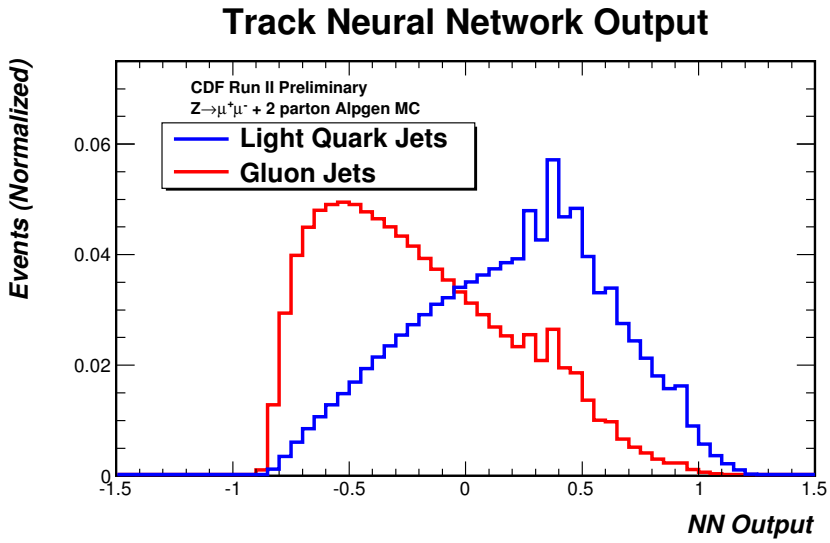

Figure 7.5: The output of the Track ANN discriminant for light-flavor quark (*blue*) and gluon (*red*) jets. We see a good level of discrimination between quark and gluon jets in this NN variable. The somewhat discrete structure comes from jets with $N_{\text{tracks}} = 2$.

7.2.3 The final quark/gluon discriminant

The purpose of the final ANN is to combine the information from the Tower and Track NN values, along with two other classes of variables: first, other, simpler variables that may offer some discrimination for quark and gluon jets; and, second, variables that may affect what the width of each jet looks like, and thus will have some bearing on how quark or gluon-like each jet looks. Variables we include in the final ANN to provide greater discrimination power include the following:

- the ratio of the sum of tower energies, ΣE, in a cone of $R = 0.4$ to that in a cone of $R = 0.7$ around the jet center, which is similar to the core energy used to help distinguish quark and gluon jets in [69];

- similarly, the ratio of the scalar sum of track momenta $\Sigma|p|$ in a cone of $R = 0.4$ to $\Sigma|p|$ in a cone of $R = 0.7$ around the jet center;

- the number of towers in cones of size R = 0.4 and R = 0.7 around the jet center, as gluon jets will typically have their energy spread across more towers;
- and, the number of tracks in cones of size R = 0.4 and R = 0.7 around the jet center, as gluon jets will typically have higher charged particle multiplicities.

Variables that we include due to their correlations with the Track and Tower NN values include the following:

- the jet E_T, as jets with higher E_T are typically more collimated and so naturally have higher Tower NN values;
- the jet η, so we avoid any dependences in the final NN on differences in the behavior of the calorimeter in different η regions;
- the jet EM fraction ($E_{\text{EM}}/E_{\text{HAD}}$), so we avoid any differences in that Tower NN value that may arise due to non-compensating aspect of the calorimeter;
- and, the number of identified interaction vertices, N_{vtx}, as events with more pileup will have more energy deposited throughout the entire calorimeter, which has the effect of making jets appear broader/less collimated.

Again, the gluon jet E_T and η distribution are reweighted to match that of the quark jets (and the N_{vtx} distributions intrinsically agree), so that the final discriminant does not show a bias towards jets based solely on these variables, but instead only with correlations of these and other variables.

The output of this final ANN is shown in Fig. 7.6 for light-flavor quark and gluon jets from the $Z \rightarrow \mu^+\mu^-$+ jets MC sample we use to test the network for overtraining. We see good discrimination between light quark and gluon jets in this final "QG" value: the former more strongly peak at a high score than they do in either the Tower or Track NN values;

meanwhile, the gluon jets have a QG value that is near -1. From these MC samples, the efficiency of a tag that rejects $\sim$ 90% of gluons would keep almost 50% of light quark jets, which is a performance almost on par with *b*-taggers. However, this performance is highly optimized for jets in MC, and as we saw in our determination of the quark and gluon jet energy scales in Sec. 5.6, there are differences in the modeling of gluon jets between MC and data. Thus, we must check that the behavior is similar in data, and calibrate the ANNs' response if it is not.

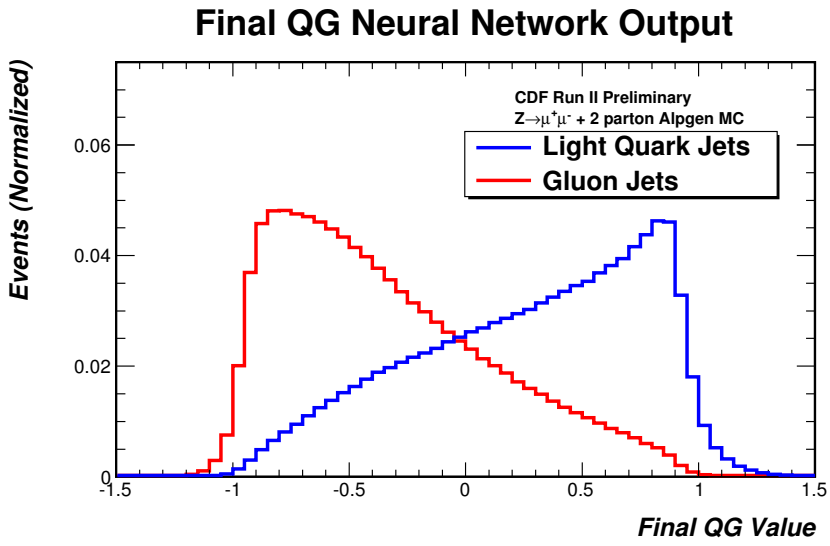

Figure 7.6: The output of the final ANN for light-flavor quark (blue) and gluon (red) jets. Higher NN scores indicate jets that are more quark-like. In MC jets, we see good separation between quark and gluon jets.

7.3 Jet QG Value Calibration

The response of the ANN quark/gluon discriminant may differ between data and MC simulation, especially since raw (uncorrected) tower energies are used in the construction of the tower ANN, as CDF does not employ individual tower energy corrections. Since our signal and most backgrounds are modeled with MC, it is necessary to calibrate the Monte

Carlo response to match the behavior of the data. We do so using an independent control region with jets similar to those in our final state, $W \to \ell\nu$ + 1 jet events, and then further validate and establish uncertainties on the modeling using data regions very similar to our diboson signal and dominant $Z + 2$ jets background regions: $t\bar{t} \to W^+bW^-\bar{b} \to \ell^\pm\nu_\ell b\bar{b}q\bar{q}'$ decays, and $W \to \ell\nu + 2$ jet events, respectively.

7.3.1 Calibration Selection Region

To form our $W + 1$ jet calibration sample, we choose data from the standard high-p_T central electron or muon triggers also used in our selection for studies of the *b*ness tagger, described in Sec. 6.4. We then apply the selection requirements summarized in Tab. 7.1. We require there be one and only one central lepton (TCE objects for electrons, or CMUP or CMX muons) with $p_T > 20$ GeV/c. To pick events consistent with a $W \to \ell\nu$ decay, we also require a large amount of missing transverse energy, $\not\!\!E_T > 25$ GeV, and a reconstructed transverse mass given by

$$m_T = \sqrt{2p_T^{\text{lepton}} \not\!\!E_T \left(1 - \cos\Delta\phi_{\ell,\not\!\!E_T}\right)},$$

where $\Delta\phi_{\ell,\not\!\!E_T}$ is the difference in ϕ in the direction of the lepton $\vec{p_T}$ and $\vec{\not\!\!E}_T$ vectors, be consistent with leptonic $W^\pm$ boson decays. To further eliminate any contributions from multijet events where jets fake our lepton $+\not\!\!E_T$ signature, we require that the $\not\!\!E_T$ not be aligned with any reconstructed jet via a cut on $\Delta\phi_{\not\!\!E_T,\text{nearest jet}}$, and that the $\not\!\!E_T$-significance be large. We also require that the events in this calibration sample have one jet with $E_T > 20$ GeV and $|\eta| < 2.0$, but have no other jets in that region with $E_T > 20$ GeV.

We consider a number of processes that may contribute to this selection, and model them with a combination of PYTHIA, ALPGEN, and MADGRAPH [73] event generators interfaced with PYTHIA for showering. The dominant contribution is $W \to \ell\nu$ production in

W + 1 jet Event Selection
$N_{\text{vtx}} > 0$
$N_{\text{leptons}} = 1$
Lepton passes TCE, CMUP, or CMX selection
Lepton $p_T > 20$ GeV/c
$\not\!\!E_T > 25$ GeV
W $m_T > 25$ GeV/c^2
$\not\!\!E_T$-significance $> 1(4)$ for $\mu(e)$ events
$\Delta\phi_{\not\!\!E_T,\text{nearest jet}} > 0.2$ rad (nearest jet with $E_T > 5$ GeV)
N_{jets} with $E_T > 20$ GeV = 1
Jet $E_T > 20$ GeV and $\lvert\eta\rvert < 2.0$

Table 7.1: Summary of event selection requirements for our W + 1 jet selection, used to calibrate parts of our QG discriminant.

association with one jet, which we model using an ALGPEN event generator interfaced with PYTHIA for showering, similar to our Z + jets MC sample. As we are largely concerned with the agreement in shapes between data and MC, we scale the MC to match the normalization of the data. Additionally, we reweight the MC to match the jet E_T and η distribution of the data, to remove these variables as a factor in any mismodeling of the other quantities used in forming the jet QG value.

7.3.2 Tower NN Calibration

In Fig. 7.7 we compare the distribution of the Tower NN value in data and MC. Our simulated model, almost entirely ALPGEN-generated W + 1 parton MC, does not match the data well, as the data appears more gluon-like. That jets in data appear more spatially spread than jets in MC is consistent with the disagreements between data and MC in jet energies from Sec. 5.6, where we found the amount of energy contained within a cone of $R = 0.4$ is higher in MC gluon jets than those in data.

We correct for these discrepancies by applying a linear shift to the Tower NN values in MC in order to match the data in the $W+1$ jet sample. We perform different linear shifts in

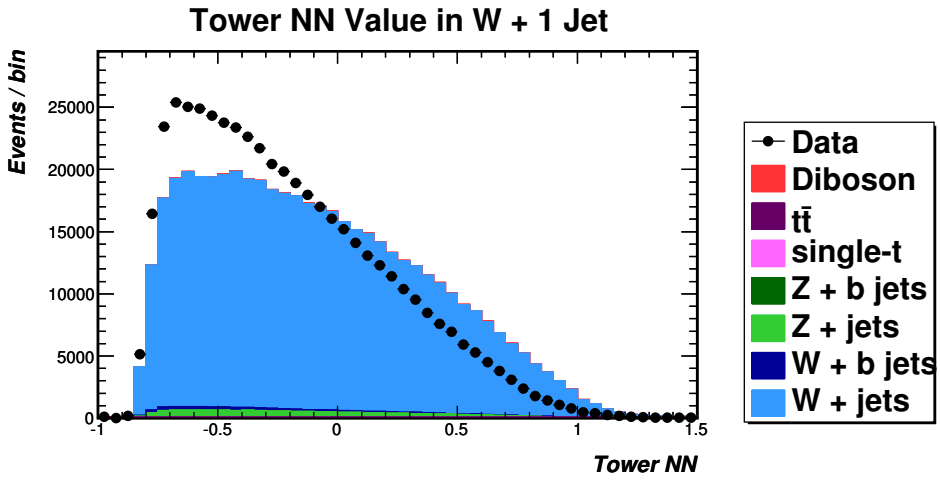

Figure 7.7: The distribution of Tower NN values in our W + 1 jet selection. We do a linear correction to the MC Tower NN values for these jets in order to improve the MC modeling of the final QG value.

four different η regions; the regions are chosen to get separate corrections for jets entirely in the central calorimeter, mostly in the central calorimeter, mostly in the plug calorimeter, or entirely in the plug calorimeter. Different corrections are also derived based on N_{vtx}, to account for possible differences in the effects of pileup between data and MC.

The corrections are made in the following way. We first make a histogram of the data jets' Tower NN values using 60 bins between -1.5 and 1.5 and a histogram of the Monte Carlo jet Tower NN output over the same range, but with 600 bins. We then compare the number of data events in each bin, starting from the lowest Tower NN values, with the number of Monte Carlo events in that same Tower NN range. If the agreement is within a specified tolerance, we move to the next data bin. If the agreement is not within that tolerance, we find the value of Tower NN output in the Monte Carlo that would serve as the best upper edge to match the number of data events in that bin, and then evenly expand or contract the MC Tower NN values to match the range of the data in that bin.

As an example, suppose that there are more events with a Tower NN between a and b in data than in Monte Carlo. We find the value $b' \in [a, b]$ for which the number of events

in the Monte Carlo with Tower NN between a and $b^{'}$ is the same as (or closest to) the number of data events with Tower NN between a and b. Then, for every jet with Tower NN between a and $b^{'}$ in the Monte Carlo, we change the Tower NN value ($\mathrm{NN_{tower}}$):

$$\mathrm{NN_{tower}} \rightarrow a + (\mathrm{NN_{tower}} - a) \times \frac{(b-a)}{(b^{'}-a)} \ .$$

Then, if there are more events with a Tower NN output between b and c in data than between $b^{'}$ and c in the Monte Carlo, we do a similar procedure, finding the value $c^{'}$ for which the number of Monte Carlo events between $b^{'}$ and $c^{'}$ is closest to the data, and changing the Tower NN output in that region accordingly:

$$\mathrm{NN_{tower}} \rightarrow b + (\mathrm{NN_{tower}} - b^{'}) \times \frac{(c-b)}{(c^{'}-b^{'})} \ .$$

The procedure is similar for the case where the number of data events in a bin is less than the number of Monte Carlo events. As an outcome of this procedure, the Tower NN values between data and MC will, by construction, agree exceptionally well.

We also ensure that correlations of the new Tower NN variable with other variables that use uncorrected tower energies are similar between data and MC. These correlations are important to model properly, as they will affect the response of the final QG value in the network. In Figure 7.8, profile histograms show the correlation between the Tower NN variable and the N_{towers} in a cone of size $R = 0.7$, and the ratio of tower energy sums: ΣE (cone $R = 0.4$) / ΣE (cone $R = 0.7$). We see that the data is not well modeled by the MC, even after correcting the Tower NN value in MC. Thus, we introduce a correction to account for this effect. For each bin (of width 0.05) in the correction Tower NN value, we make a shift of $N_{\mathrm{towers}} \rightarrow a * N_{\mathrm{towers}} + b$, where a and b are determined using the profile histogram. We need only a shift of the form ΣE ($R = 0.4$) / ΣE ($R = 0.7$) $\rightarrow a * \Sigma E$ ($R = 0.4$) / ΣE ($R = 0.7$) for the energy ratio in order to better match the data. These

corrected values are determined separately for forward and central jets, as well as for events with different N_{vtx}.

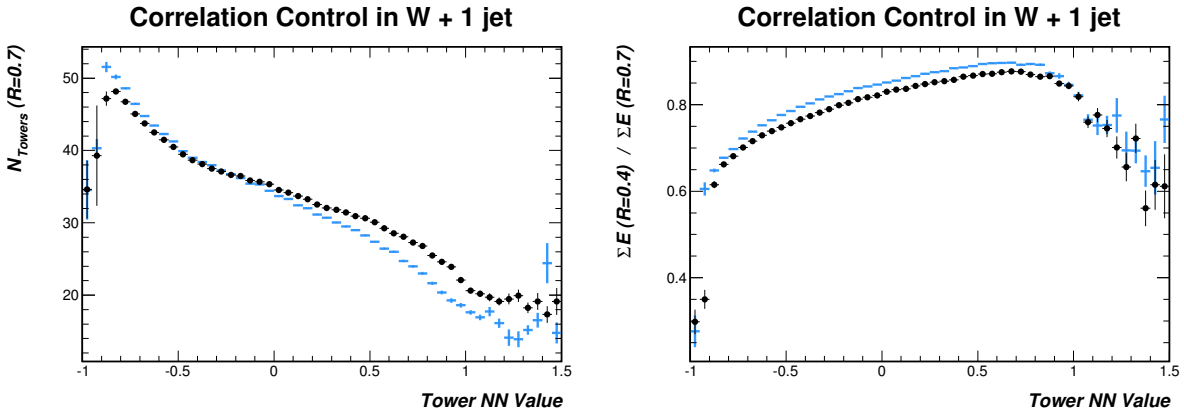

Figure 7.8: Profile histograms showing the relationship between the Tower NN value (the corrected one in MC) and the N_{towers} in a cone of size 0.7 (left), and the ratio of ΣE (cone 0.4) / ΣE (cone 0.7) (right), comparing data (black) with MC (light blue).

7.3.3 Track NN Calibration

We also check the output of the Track NN values and also see a difference between data and MC. Some of this difference can be attributed to differences between the number of tracks in jets in data and MC, due to an overefficiency of track reconstruction in the MC. We reweight the MC events to match the distribution of number of tracks in the data, and the result for jets with more than two tracks is shown in Figure 7.9. We see some slight disagreement between the MC and data in the Track NN output, but it is much smaller in degree than in the Tower NN case, due to the better modeling of tracks. The data Track NN values, compared to those of the MC, tend to be slightly higher (more quark-like) in jets with more than two tracks within a cone of $R = 0.7$. To correct for this discrepancy, we input a small, constant linear shift to the Track NN values. Because the response function

of the neural network is tanh(), we apply a correction to MC as

$$\mathrm{NN}_{\mathrm{track}} \rightarrow \tanh[\tanh^{-1}(\mathrm{NN}_{\mathrm{track}}) + c] \, ,$$

where we determine the value for c by comparing the mean of the data and MC distributions. This correction is done separately for events with different N_{vtx}, but the differences between events with different N_{vtx} are, on the whole, rather small.

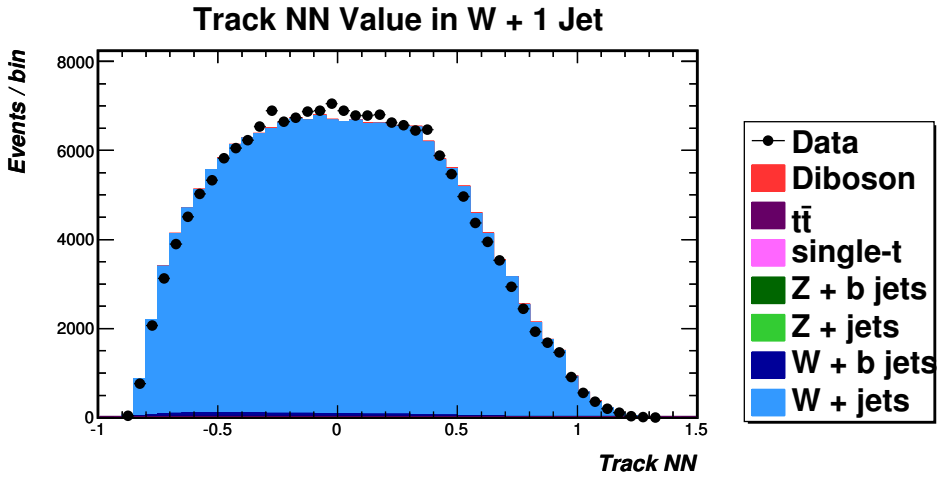

Figure 7.9: The distribution of Track NN values in our W + 1 jet selection for jets with three or more tracks within a cone of R = 0.7. After reweighting the N_{track} distribution we see good agreement between data and MC. We do a minor correction to the MC Track NN values for jets with three or more tracks based on the differences in these distributions.

7.3.4 QG Value Calibration Results

All calibrated variables in MC are input directly into the final ANN without retraining the network. The effect of the calibrations on the final jet QG value is shown via a comparison of data and MC in Z + 1 jet events, using the selection from Sec. 4.4, in Fig. 7.10. This sample is orthogonal to the calibration region. We see much better modeling of the QG values after the calibrations, though we do see some disagreements, especially in

a region at lower QG value. These remaining differences will be evaluated and handled separately in the following section.

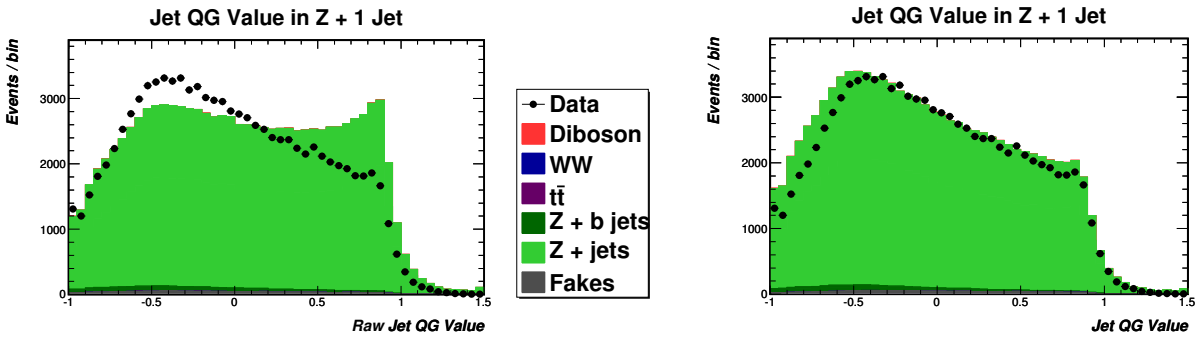

Figure 7.10: The distribution of the final QG neural network values in our Z + 1 jet selection. We show the MC distribution before (left) and after (right) the calibrations of the QG value response. We see much better agreement between data and MC after corrections.

7.4 Evaluation of "Light Tag" Efficiency and Mistag Rate

Similar to our characterization of the jet *b*ness tagger, described in Secs. 6.3-6.6, we determine a mistag rate and tag efficiency for placing a cut on the jet QG value in order to form a light-flavor tag region in our data selection. The procedure follows that described in detail for the jet *b*ness: using two sets of data with differing light quark content, we iteratively determine a tag and mistag rate for a cut on the jet QG value in both data and MC. We can then correct the MC to match the behavior of the jet QG cut in the data.

For a sample rich in light quarks, we again use a $t\bar{t} \to W^+bW^-\bar{b} \to b\bar{b}\ell^{\pm}\nu_{\ell}q\bar{q}'$ selection where we now select the two jets with lowest *b*ness values to pick out the two jets most likely from the hadronic decay of the W boson. We use a W + 2 jet sample as a sample richer in mistags, and similar in quark/gluon content to our Z + jets background. Tab. 7.2 summarizes the cuts placed to form the two data samples: the W + 2 jet sample is similar to the previously described W + 1 jet sample, except we modify the cuts on the

W + 2 jets selection		$t\bar{t}$ selection
	$N_{vert} > 0$	
	$N_{\text{leptons}} = 1$	
	Lepton passes TCE, CMUP, or CMX selection	
	Lepton $p_T > 20$ GeV/c	
	$\not\!\!E_T > 25$ GeV	
$\Delta\phi_{\not\!\!E_T,\text{nearest jet}^*} > 0.4$ rad		$\Delta\phi_{\not\!\!E_T,\text{nearest jet}^*} > 0.2$ rad
$\not\!\!E_T$-sig $> 4^\dagger$		
W $m_T > 25$ GeV/$c^{2\dagger}$		
		Total sum $E_T > 300$ GeV
N_{jets} ($E_T > 20$ GeV) ≥ 2		N_{jets} ($E_T > 20$ GeV) ≥ 4
		2$^{\text{nd}}$ highest $b$ness jet > -0.5
		Two highest $b$ness jets $E_T > 20$ GeV
1$^{\text{st}}$/2$^{\text{nd}}$ jet $E_T > 25$ GeV		Two lowest $b$ness jets $E_T > 25$ GeV
	Jets' $\lvert\eta\rvert < 2.0$	
	ΔR between jets > 0.7	

*The nearest jet here is the nearest cluster with raw $E_T > 5$ GeV.
†Only applied to events with a central electron.

Table 7.2: Summary of event selection requirements for our $t\bar{t}$ lepton + jets selection and our W + 2 jets selection, used to understand the modeling of events in our QG discriminant. Cuts in the center are shared cuts in the two samples. The total sum E_T is defined as the sum of the lepton p_T, $\not\!\!E_T$, and E_T of all jets with $E_T > 15$ GeV.

jets to better match those used in the signal region of our $ZW/ZZ \to \ell^+\ell^- jj$ search; meanwhile, the $t\bar{t}$ selection eschews the $\not\!\!E_T$-sig and m_T cuts to reduce multijet backgrounds in favor of requiring a minimum scalar sum of the E_T of identified objects (jets, $\not\!\!E_T$, and the lepton p_T) in the event.

	W + jets Selection	$t\bar{t}$ Selection
W + jets*	21520 ± 2150	38.7 ± 3.9
W + b jets*	937 ± 375	13.8 ± 5.5
Z/DY + jets	1249 ± 125	3.1 ± 0.3
Z/DY + b jets	86 ± 34	1.4 ± 0.6
WW/WZ	1386 ± 83	5.9 ± 0.4
single t	767 ± 77	19.6 ± 2.0
$t\bar{t}$	1378 ± 83	469 ± 28
$t\bar{t}$ (b-jets)		108 ± 7
$t\bar{t}$ (q-jets)		361 ± 22
Total Expected	27319*	551 ± 30
Data	27319	579

* W+ jets samples have been scaled to as to produce agreement with data in the number of events.

Table 7.3: The number of events in the W + 2 jets and $t\bar{t}$ lepton + jets region, showing only the uncertainties assigned on the normalization of each sample. The distinction between b and q jets in the $t\bar{t}$ sample refers to the lower two $b$ness jets: events where both jets are matched to non-b quark jets are labeled "q-jets", while if one of the jets is matched to a b jet, it is labeled "b-jets".

Because we are looking for jet QG shape differences between data and MC that will translate to acceptance uncertainties when we place a cut on the jet QG value, we scale the number of W+ jet events in the MC to match the data in our W + 2 jets sample, and apply that same scaling to W+ jet samples in our modeling of the $t\bar{t}$ selection. The number of events in each sample is shown in Tab. 7.3. We take uncertainties on the normalizations of our $t\bar{t}$, single-t, WW/WZ, W/Z + jets, and W/Z + b jets samples to be 6%, 10%, 6%, 10%, and 40%, respectively, based on uncertainties from the theoretical cross sections (for the top quark and diboson samples), or as the magnitude of the necessary scaling in order to agree with data (for the W/Z+ jets samples).

The distributions of the maximum and minimum QG values of the two jets considered are shown in Fig. 7.11. We see fairly good modeling in the $t\bar{t}$ sample, especially when taking into consideration the small normalization difference between the data and MC model here. We see poorer modeling in the $W+2$ jet sample, where, after our calibrations, jets in MC are slightly more gluon-like than those in data. The reasons for this mismodeling may be due to residual differences between data and MC in the QG response after the applied calibrations, differences in the quark/gluon content between data and MC, and differences in other related variables (like the jet E_T spectrum) that may be correlated with differences in the QG value.

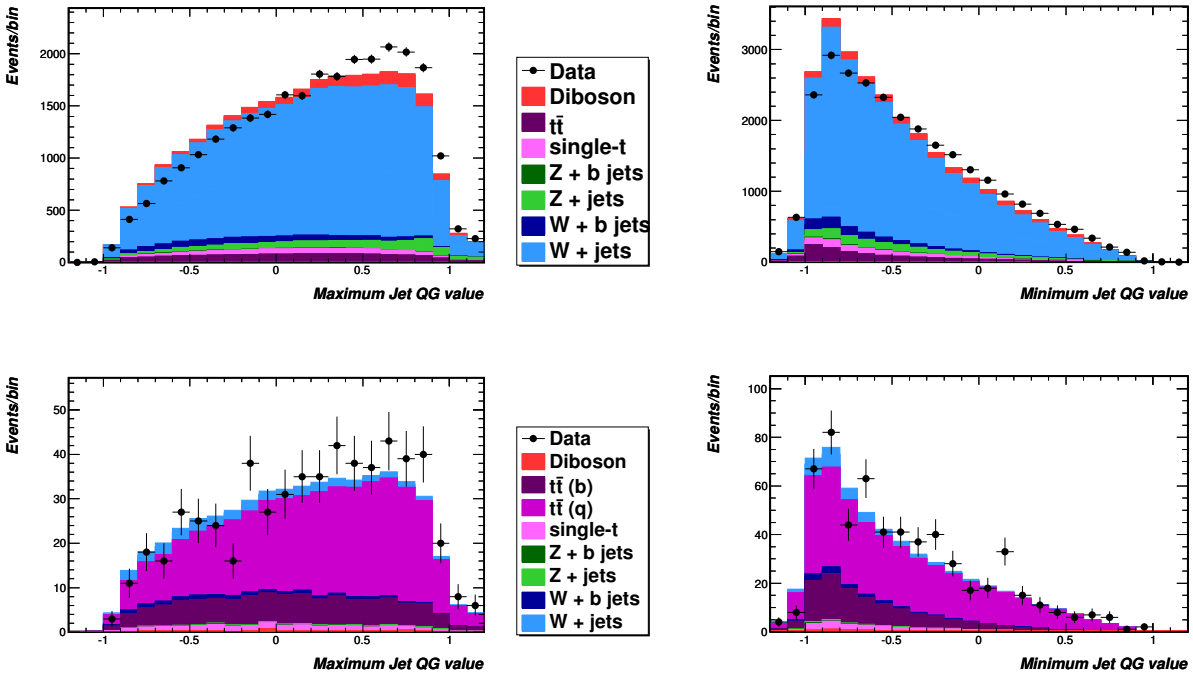

Figure 7.11: Distribution of the maximum (*left*) and minimum (*right*) jet QG values of the two jets in our $W+2$ jet (*top*) and $t\bar{t}$ (*bottom*) samples. The distinction between *b* and *q* jets in the $t\bar{t}$ sample refers to the lower two *b*ness jets: events where both jets are matched to non-*b* quark jets are labeled "*q*-jets", while if one of the jets is matched to a *b* jet, it is labeled "*b*-jets".

We can determine an efficiency for quarks to pass a certain jet QG value cut and

gluons to be "mistagged" using this cut with the $t\bar{t}$ and $W+2$ jet samples in a way directly mirroring the jet *b*ness efficiency and mistag rate determined in a $t\bar{t}$ and $Z+1$ jet sample. The efficiency measured in data – $e_D(q)$, as it is a function of the QG cut placed – may be expressed as:

$$e_D(q) = \frac{e_{raw}(q) - s_m(q)m_{MC}(q)f_g}{1 - f_g} \ , \tag{7.1}$$

where e_{raw} is the fraction of data events passing the QG cut; m_{MC} is the mistag rate for gluons to pass the cut, as measured in MC; s_m is a scale factor on the mistag rate in MC to match the mistag rate measured in data; and, f_g is the fraction of gluon jets in the sample. We can write a similar expression for getting the mistag rate from

$$m_D(q) = \frac{m_{raw}(q) - s_e(q)e_{MC}(q)f_q}{1 - f_q} \ , \tag{7.2}$$

where m_{raw} is the fraction of data events passing the QG cut; e_{MC} is the efficiency for quarks to pass the cut, as measured in MC; s_e is a scale factor on the efficiency in MC to match the mistag rate measured in data; and, f_q is the fraction of quark jets in the sample. Uncertainties on these quantiites follow those shown in Eqs. 6.7 and 6.8.

We measure the QG-tag efficiency in the $t\bar{t}$ sample, where we have a very small number of gluon jets, and measure the mistag rate in the $W+2$ jets sample, where the gluon fraction is much larger, and similar to our $Z+2$ jet signal region. Tab. 7.4 shows the efficiency and mistag rate for our given cut at minimum QG > 0.0, measured in both data and MC. For this operating point, we see that the MC underestimates the rate for quark jets to pass the jet QG cut, while more correctly modeling the observed mistag rate.

Quantity	QG Cut	Data	MC	% Difference	% Error
Mistag Rate	> 0.0	0.087	0.088	-1%	31%
Tag Efficiency	> 0.0	0.295	0.241	+18%	12%

Table 7.4: The efficiency and mistag rates for a cut on the minimum jet QG value in our two-jet samples at > 0.0, determined from comparisons of data and MC in our $W + 2$ jet and $t\bar{t}$ control regions.

CHAPTER 8

SIGNAL EXTRACTION AND RESULTS

While there are a number of characteristics that distinguish the ZW/ZZ diboson signal from our dominant Z+ jets background, the most prominent is the reconstructed invariant mass of the two jets in our selection. Shown in Fig. 4.7, our signal has a Gaussian-like peak around 80 GeV/c^2. The jet energy resolution is not fine enough to distinguish between the mass of the W and Z bosons, thus both $W \to q\bar{q}'$ and $Z \to q\bar{q}$ decays make up that peak. Meanwhile, the Z+ jets background has a significantly lower dijet mass spectrum that, after a peak at about 50 GeV/c^2, is smoothly falling in the region where our signal peaks.

The difference between signal and background in the m_{jj} spectrum motivates a strategy to extract our diboson signal via a fit to this distribution. We perform the fit to the dijet mass spectrum between 20 GeV/c^2 and 260 GeV/c^2. In order to increase our sensitivity, we split up the Z + 2 jet selection using our tagging variables and obtain regions of data with larger S/B. We place a cut on the jet *b*ness variable (Chap. 6) to increase our sensitivity by picking out $ZZ \to \ell^+\ell^-$ + $b\bar{b}$ events, as the bulk of our background does not contain b jets. We also place a cut on the jet QG variable (Chap. 7) since the remaining signal events should contain two light-quark jets from the decay of a W or Z boson. Because heavy-flavor jets tend to be more spatially spread than light-flavor jets, it is important that we apply the b-tagging selection before applying a light quark-tagging cut. Thus, we separate our dijet mass fit into three separate channels:

1. A heavy-flavor tagged region, using a cut on the jet *b*ness values.

2. A light-flavor tagged region composed of events that fail the *b*-tagging requirement, but pass some cut on the jet QG values.

3. A no-tag region composed of events that fail both cuts.

The inclusion of a no-tag region allows for the collection of quark jets that end up failing the two tagged requirements; all events with 20 GeV/c^2 $\leq m_{jj} \leq$ 260 GeV/c^2 and pass the selection requirements in Tab. 4.1 are included in the signal extraction in order to maximize our acceptance. The structure of events into the three fitting channels is shown in Figure 8.1.

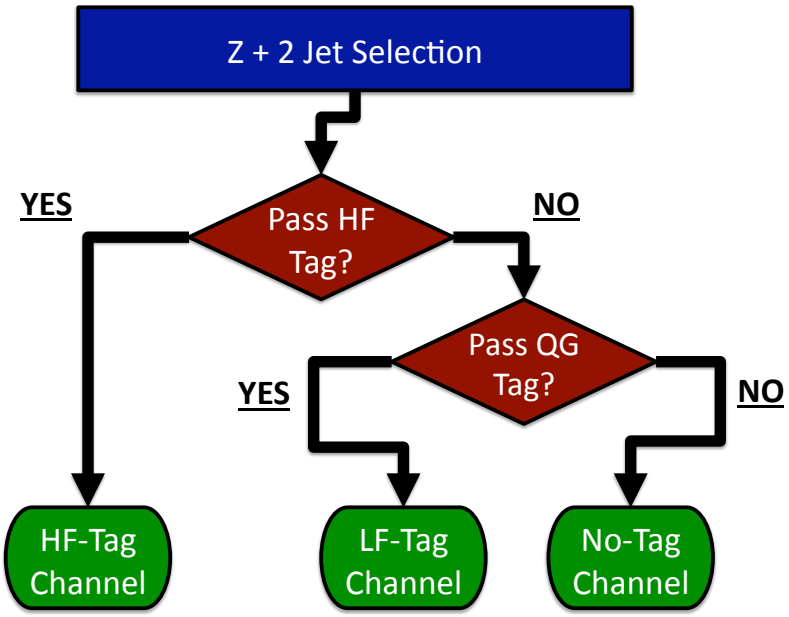

Figure 8.1: Flow chart showing how the events in our Z + 2 jet signal region are divided into our three fitting channels. First, we use the jet *b*ness values to tag jets likely from a $Z \rightarrow$ heavy flavor quarks decay. For events that fail that cut, we then use the jet QG values to tag jets likely from a $Z \rightarrow$ light flavor quarks decay. The remaining events enter a no-tag channel.

We determine where to place our tagging cuts by maximizing $S/\sqrt{B}$ in a region under

the dijet mass peak of the signal (between 60 GeV/c^2 and 120 GeV/c^2). We find the best placement for our heavy-flavor tag cut is to require the minimum jet *b*ness of the two leading jet to be greater than 0.0. Similarly, we find the best definition of our light-flavor tag channel to be events with minimum jet QG value greater than 0.0. Both minima are rather shallow, and cuts over a range around these threshold values would yield similar sensitivity. The two tagging variables are shown in Figures 8.2 and 8.3. The agreement in the jet *b*ness is rather good, while the modeling of the jet QG value suffers from a disagreement between data and MC that is similar to that in the W + 2 jet region. Our corrections on this variable derived from that region, along with the calculation of the systematic uncertainties on that correction, should hold in the signal region.

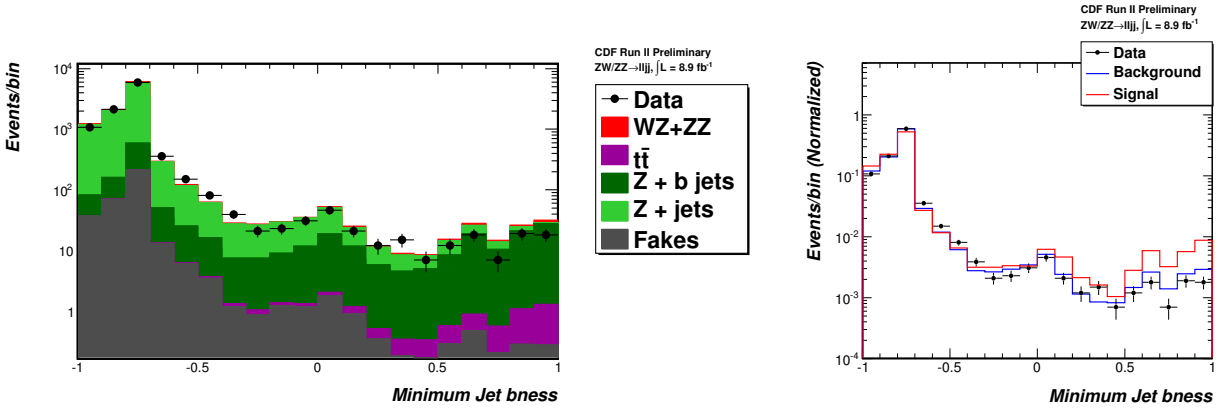

Figure 8.2: The distribution of the minimum jet *b*ness in our Z + 2 jet region. Overall we see pretty good agreement in data and MC, and we see MC samples with *b*-quarks in them ($t\bar{t}$ and $Z + b\bar{b}$) show high values, as expected.

The total number of events and the events in each channel of the dijet mass fit are shown in Tab. 8.1. The number of events in the data and predicted by our background model are in good agreement within the systematic uncertainties, described in further detail in Sec. 8.2. The distributions of the dijet mass in each of the fitting channels are shown in Figs. 8.4-8.6. Other variables describing the kinematics of the jets are shown in App. B.

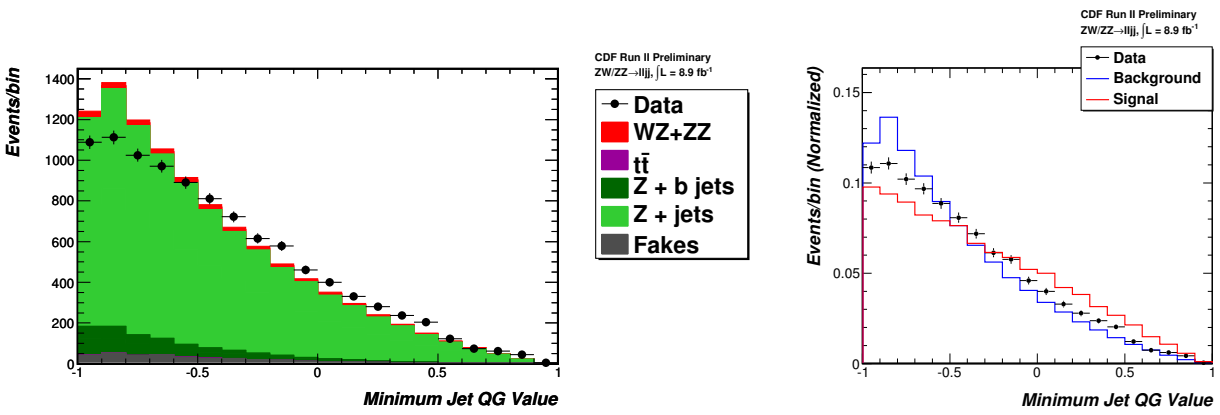

Figure 8.3: The distribution of the minimum jet QG values in our Z + 2 jet region. Our signal has a higher jet QG value, as expected since it is composed largely of quark jets.

	All Events	HF-Tag Chan.	LF-Tag Chan.	No-Tag Chan.
Z + jets	$8,667 \pm 1,113$	93 ± 14	$1,454 \pm 307$	$6,721 \pm 968$
Z + b jets	714 ± 299	111 ± 48	53.8 ± 25.5	536 ± 230
$t\bar{t}$	9.2 ± 0.9	3.3 ± 0.4	0.7 ± 0.1	5.2 ± 0.6
Fakes	330 ± 165	4.8 ± 2.4	39.4 ± 20.3	283 ± 142
Total Bkg.	$9,720 \pm 1,247$	212 ± 55	$1,617 \pm 325$	$7,890 \pm 1,071$
ZW + ZZ	313 ± 29	12.8 ± 1.6	84.8 ± 12.3	205 ± 22
Predicted Events	$10,033 \pm 1,259$	225 ± 55	$1,706 \pm 331$	$8,102 \pm 1,080$
Data Events	9,846	172	1,724	7,950

Table 8.1: The expected number of events (compared with data) in each channel of the fit of the dijet mass distribution. The uncertainties here represent all systematic uncertainties, including the luminosity. While we let them remain unconstrained in the fit, we include 10% and 6% uncertainties on the normalization of the Z+jets and signal samples, respectively, in this table.

8.1 Details of Fitting Procedure

After splitting the data into the three tagging categories, we extract the number of signal events using a χ^2-minimization fit to data. The fit is done using many aspects of the `mclimit` limit-setting program, described in [74]. We supply histogram templates for our signal and background samples, and use those templates to fit data or pseudo-data

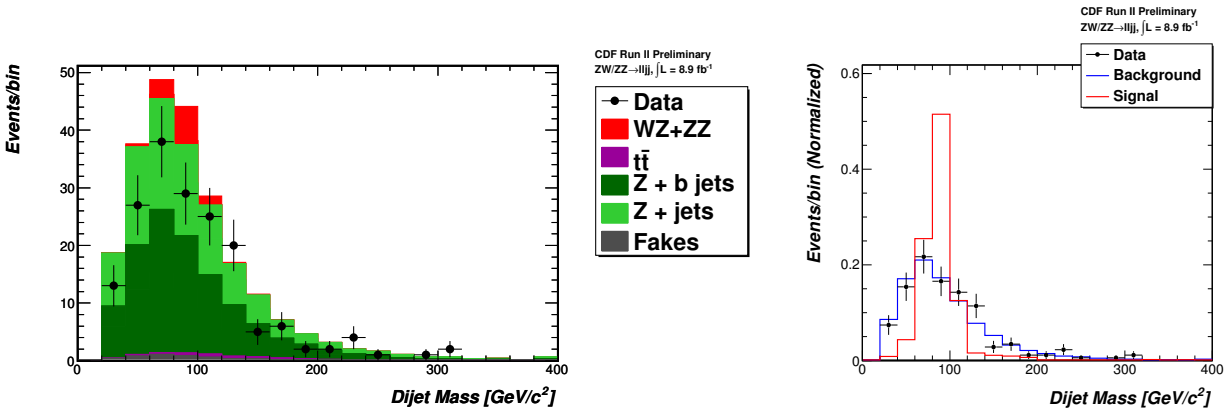

Figure 8.4: The dijet mass distribution in our heavy-flavor-tagged channel, with MC normalized absolutely (left) and to equal area with the data (right).

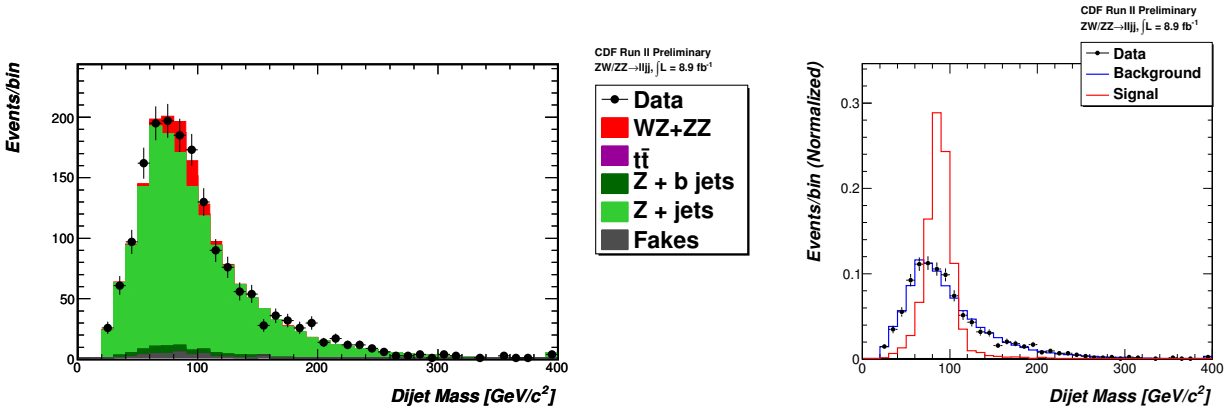

Figure 8.5: The dijet mass distribution in our light-flavor-tagged channel, with MC normalized absolutely (left) and to equal area with the data (right).

histograms. The bins of pseudo-data histograms are filled by throwing random numbers following Poisson distribution with a mean equal to the expected number of events in that bin. We also provide a list of systematic uncertainties that may change the normalization and/or the shapes of the templates. These systematic uncertainties, described in Sec. 8.2, are allowed to vary in order to minimize the χ^2 of the fit, though they are typically under a Gaussian constraint so as not to stray far from their nominal values. Therefore, these "nuisance parameters" are fit for in the procedure. The systematic uncertainties are also allowed to vary, according to their Gaussian constraints, in generating the pseudo-

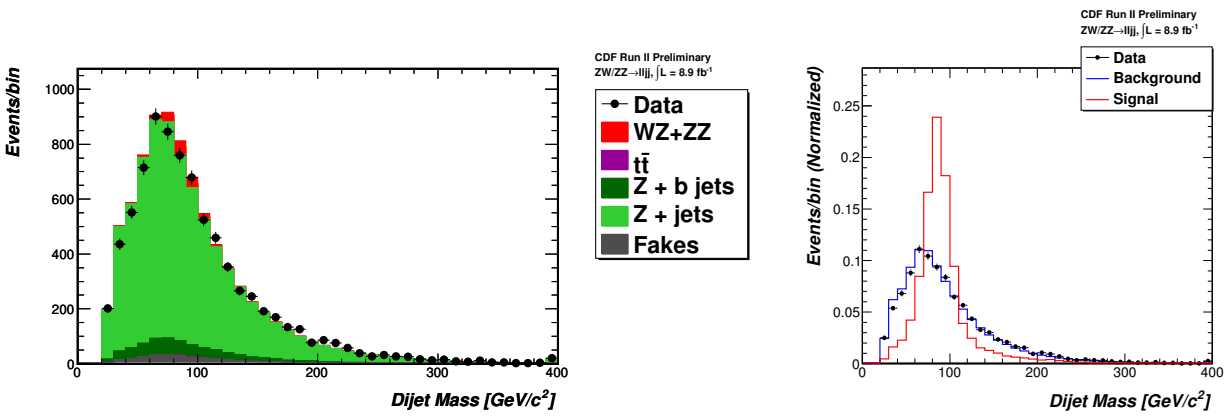

Figure 8.6: The dijet mass distribution in our no-tag channel, with MC normalized absolutely (left) and to equal area with the data (right).

data.

We use five templates corresponding to the four main components of our background modeling plus our signal model to fit the data: Z+ jets, $Z+b$ jets, $t\bar{t}$, fakes, and our ZW/ZZ diboson signal. The normalization of each template comes from the method described in Sec. 4.3, but in the fit the normalization of some templates is allowed to vary without any constraints. Since our goal is to measure the cross section for ZW/ZZ production in this leptons plus jets channel, we allow the normalization of the signal template to vary unconstrained. Additionally, we allow the normalization of our the Z + jets background to also vary unconstrained, as this normalization should be largely constrained by the fit to regions with little signal (low and high dijet mass bins). The normalization of Z + b jets template is also unconstrained, but is pinned to move in the same direction as the Z + jets normalization, preserving the ratio of $\sigma_{Z+b \text{ jets}}/\sigma_{Z+\text{jets}}$ as measured in [51].

The fit is performed simultaneously in the three fitting channels, with all normalizations and other nuisance systematic parameters correlated among the three channels. A χ^2 parameter is constructed following a calculation of a binned likelihood, described in greater detail in [75]. For each set of pseudo-data (and for the actual data), we first perform a "null hypothesis" fit, where we assume there is no signal, calculating a χ^2_{null}. Then we per-

form a fit using the "test hypothesis", where we include in the fit the possibility of signal, extracting a χ^2_{test}. We can then calculate a $\Delta\chi^2 = \chi^2_{\text{test}} - \chi^2_{\text{null}}$, which reflects how signal-like the data or pseudo-data is. Note that $\Delta\chi^2$ is always less than zero, since it is always the case that $\chi^2_{\text{test}} < \chi^2_{\text{null}}$, a reflection of the fact that we can always perform a better fit with an additional free parameter with which to fit.

The final result of the fit is the minimum χ^2_{test}, χ^2_{null}, $\Delta\chi^2$, and a list of the normalization and nuisance parameters used to create the best fit. The most important parameter to extract in this analysis is the normalization of the ZW/ZZ diboson template (which is highly correlated with $\Delta\chi^2$), as that may then be translated to a measured cross section.

8.2 Systematic Uncertainties

In performing the fit to data, we simultaneously fit for systematic errors that can affect the shape or normalization of the dijet mass distribution. Normalization uncertainties are handled by assigning what a 1σ variation in a systematic parameter corresponds to, and then determining how much that 1σ variation changes the normalization of a specific template. The systematic parameters are then treated as fitting parameters in the χ^2 minimization, but are typically assigned a Gaussian constraint, so that deviations from their nominal (0σ) value incurs a penalty in the calculation of the χ^2.

Uncertainties that may change the shape of the dijet mass distribution in our background models are handled through a histogram morphing procedure, described in greater detail in [76]. The goal of this morphing procedure is to produce a m_{jj} template given a desired variation of a particular systematic parameter. We use a simple "vertical" morphing procedure where the number of events in each bin vary linearly with the distance in σ from the expected value of the systematic parameter. We obtain dijet mass shapes corresponding to $+/-1\sigma$ variations in a systematic parameter by varying that parameter,

obtaining new background templates for the m_{jj} distribution, and using these templates along with our nominal background template as endpoints in the vertical morphing.

Unfortunately, simply performing the morphing procedure with these templates may lead to two problems. When performing the morphing procedure between the -1σ and $+1\sigma$ template, it is not guaranteed that the produced 0σ template will agree with the nominal background template, as it should. Alternatively, if we perform the morphing procedure between -1σ and 0σ for desired variations $< 0\sigma$, and between 0σ and $+1\sigma$ for variations $> 0\sigma$, there may be a discontinuity in the rate of change of the templates at 0σ that may produce a false minimum in our fitter. To avoid these issues, we symmetrize the error in each bin of the m_{jj} distribution by comparing the error from $+1\sigma$ and -1σ variations with the nominal background template. The histograms are then morphed using these symmetrized templates, where we are ensured that we get the correct central-value template with no discontinuities in the rate of change.

To handle shape systematics in our signal samples, we use a more analytic procedure, as the histogram morphing procedure described above does not always exhibit smooth variation of the mean and width of a Gaussian-like distribution. We fit the central distribution in each channel to a Gaussian function on top of a 4^{th} degree polynomial, and then fit $+1\sigma$ and -1σ template shapes with that same function, only allowing the Gaussian parameters to change. Then, we fit the resulting fit parameters to a line, forcing the lines to pass through the central value parameters, and draw intermediate shapes using those fits. This avoids using the histogram morphing in `mclimit`, which can give trouble in the χ^2 minimization for these samples. These fits are shown in App. D.

We now give a summary of the systematic uncertainties considered in this analysis. A summary of the uncertainties considered in the fit is shown in Tab. 8.3.

8.2.1 Cross Section/Normalization Uncertainties

A summary of the independent normalization uncertainties assessed on the background templates are given below:

- Z+ jets — as mentioned in Sec. 8.1, the normalization of the Z+ jets template is not constrained by any systematic uncertainty, due to the way in which the normalization is determined in the Z + 1 jet comparisons between data and MC;
- Z + b jets — while the Z + b jets template's normalization is pinned to move with the unconstrained normalization of the Z+ jets template, an additional uncertainty on the relative Z + b jets cross section of 40% (see [51]) is applied;
- $t\bar{t}$ — we take an uncertainty of 6.5% on the $t\bar{t}$ template's normalization based on the theoretical uncertainty on the production cross section;
- Fakes — as discussed in Sec. 4.3.2, we take a 50% uncertainty on the fake rate for jets to fake electrons (where we assume all fake rates for the various electron definitions and jet energies are correlated) based on the fake rates derived from different jet triggers;

8.2.2 Jet Energy Scale and Resolution Uncertainties

A significant uncertainty in this analysis is in the proper modeling of the jet energy scale (JES) by our Monte Carlo simulations. The details of the jet energy scale uncertainty are given in Sec. 5.6.4. As stated there, the uncertainty on the difference between quark jet energy scale in data and MC is $\approx$ 2.7% of the jet energy, while the uncertainty on the gluon jet energy scale is $\approx$ 4.4%. These uncertainties are anti-correlated, and so we obtain "+1σ_{JES}" templates by moving the quark JES up by 2.7% while moving the gluon

JES down by 4.4%; similarly, the $-1\sigma_{JES}$ templates are created with shifts of -2.7% and +4.4% in the quark and gluon JES, respectively.

The effects of the JES uncertainty on template normalizations are considered for all templates derived from MC (Z+ jets, $Z + b$ jets, $t\bar{t}$, and ZW/ZZ), while correlated shape uncertainties are also considered for all but the $t\bar{t}$ template, whose contribution is small enough that shape uncertainties don't affect the result of the fit.

To evaluate a $+1\sigma$ uncertainty template due to the jet resolution, we smear the jet energies MC samples using a Gaussian distribution with mean of 1.0 and width $0.03 \cdot E_T + 1.7(\text{GeV})/E_T$, where E_T is in units of GeV. From this, we get rate systematics to all MC templates, and shape systematics for everything but $t\bar{t}$. We take the smeared template as the $+1\sigma$ systematic template shape, and interpolate/extrapolate for all nuisance parameters using this shape and the central one. We allow the resolution to get "better" by extrapolating to negative values, in units of σ.

8.2.3 Lepton Energy Scale and Resolution Uncertainties

We raise and lower the energy/momentum scale of all leptons in MC samples by $\pm 1\%$, and see a slight difference in acceptance, that we take as an uncertainty on the rate. No differences in the shape of the templates are noticeable, and so we do not include a shape uncertainty for the lepton energy scale. As an uncertainty on the lepton energy/momentum resolution, we smear all lepton energy/momenta using a Gaussian distribution with a width of 2% of the lepton's energy. This is used to establish a rate uncertainty for changing the lepton energy resolution; no significant change in the shape of the dijet mass templates is observed.

8.2.4 Jet bness and Jet QG Value Cut Uncertainties

Studies of the differences between the mistag rate and tag efficiency in data and MC are summarized in Secs. 6.5 and 6.6. Our cut requiring the minimum jet *b*ness to be greater than 0.0 in order to enter our heavy-flavor tag channel has a higher mistag rate in data than predicted by MC simulation, and a lower efficiency for selecting *b* jets. Rather than apply a simple scale factor to modify the MC to make this change, we implement corrections to the MC by modifying the location of the *b*ness cut in MC simulated samples in order to match the mistag rate/efficiency seen in the data. Thus, the cut for non-*b* jets is lowered to *b*ness > -0.25, and the cut for *b* jets is raised to *b*ness $> +0.35$. These changes in the cuts are summarized in Tab. 8.2.

To evaluate a systematic on these efficiency and mistag rate values, we raise and lower the *b*ness cuts in MC in a correlated fashion to match the $\pm 1\sigma$ values on the *b*-tag efficiency and mistag rate given in Tab. 6.4. These changes lead to rate uncertainties that affect all MC templates, not just the heavy-flavor tagged ones, as events can move across channels depending on where the *b*ness cut is. While we see no variation in shape for the signal and $t\bar{t}$ template, we do consider a correlated shape systematic for the Z + jets and Z + *b* jets templates due to the variation in the *b*ness cut.

Similar to the treatment of the *b*ness cut, we use our studies of the efficiency of a jet QG value cut in a $t\bar{t}$ signal-like region and W + 2 jets background-like region (see Sec. 7.3) to modify the cut on MC samples to better match the data. We also modify the cut to establish $\pm 1\sigma$ templates for our MC samples. We obtain uncertainties on the normalization of MC templates in the light-flavor tag and untagged channel (note they will be anti-correlated), along with shape uncertainties for the Z+ jets and $Z + b$ jets templates due to variations in this cut. The changes made in the cuts on MC samples are also summarized in Tab. 8.2.

	Data	MC	Equivalent MC Cuts (-1σ,0σ,$+1\sigma$)	Applied To ...
Jet *b*ness cut				
Mistag Rate	0.0819 ± 0.0034	0.0720	$(-0.27, -0.24, -0.18)$	jets not matched to *b* quarks
Tag Efficiency	0.622 ± 0.049	0.684	$(+0.1, +0.35, +0.6)$	jets matched to *b* quarks
Jet QG value cut				
Signal Efficiency	0.295 ± 0.034	0.241	$(-0.0325, -0.09, -0.14)$	jets matched to quarks
Background Efficiency	0.087 ± 0.027	0.088	$(0.09, -0.0175, -0.11)$	jets matched to gluons

Table 8.2: The efficiency and mistag rates for our *b*ness and QG value cuts evaluated in data and MC, along with the necessary cut value changes in MC to model the proper rates and the uncertainties on them.

8.2.5 Uncertainties from α_s

ALPGEN uses the following expression for calculating the strong coupling constant, α_s [49]:

$$\alpha_s(Q) = \frac{1}{b_5 \log\left(Q^2/\Lambda_5^2\right)} - \frac{b_5'}{(b_5 \log^2\left(Q^2/\Lambda_5^2\right)) \log\log\left(Q^2/\Lambda_5^2\right)}$$

where the constants b_5 and b_5' are the 1- and 2- loop coefficients of the QCD β function, respectively, for 5 effective quark flavors, with some cutoff energy scale Λ_5. The strength is determined by the parameter Q, called the factorization and renormalization scale. Changes in Q will correspond to changes in the coupling strength of the strong force, which may change the amounts and spectrum of initial and final state radiation, along with some of the characteristics of hard partons produced in association with the Z boson. In terms of the dijet mass spectrum of our Z + jets sample, as we increase the choice of Q, we expect a narrower dijet mass spectrum.

In generating our ALPGEN Z+ jets and $Z+b$ jets samples, we use a typical scale choice of

$$Q_{0\sigma}^2 = m_Z^2 + p_{T,Z}^2$$

which reflects the energy of the Z+ jets system. We asses a systematic on this factorization/renormalization scale by generating samples with higher and lower Q^2, increasing or decreasing the Q^2 by a factor of 2. We assign the $Q^2 = 2Q_0^2$ to the $+1\sigma$ template, and the $Q^2 = \frac{1}{2}Q_0^2$ to the -1σ template. This systematic uncertainty is applied only as an uncertainty on the shape of the Z+ jets and $Z + b$ jets m_{jj} templates.

Our signal sample is generated via PYTHIA, for which we may modify different but similar parameters in the generation that govern the strength and evolution of α_s at different scales in order to increase or decrease the amount of initial and final state radiation ("ISR" and "FSR", respectively). We use MC generated with more and less ISR and FSR

to establish differences in the shape of our ZW/ZZ templates—we do not see any significant changes in the normalization of the templates. In the fit, we allow these templates to move due to changes in the ISR/FSR in a way that is independent of changes in the Q^2 for the Z+ jets samples; that is, the ISR/FSR in our signal MC and the factorization/renormalization scale in our Z+ jets backgrounds are treated as uncorrelated.

Due to the smallness of the size of the sample, we do not include any systematics due to differences in ISR/FSR on the $t\bar{t}$ template.

Systematics in Fit	channel	WZ and ZZ	Z + jets	Z + b jets	$t\bar{t}$	Fakes
σ/Norm.	all	*unconstr.*	*unconstr.*	$\pm$40%	$\pm$6.5%	$\pm$50%
Jet Resolution	HF-Tag	$\pm$0.8%	$\pm$0.3%	$\pm$1.0%	$\pm$0.2%	
	LF-Tag	$\pm$1.0%	$\pm$0.7%	$\pm$1.5%	$\pm$6.2%	
	No-Tag	$\pm$0.6%	$\pm$0.9%	$\pm$0.7%	$\pm$1.1%	
Jet Energy Scale	HF-Tag	$\pm$4.0%	$\pm$4.8%	$\pm$3.8%	$\pm$4.0%	
	LF-Tag	$\pm$1.5%	$\pm$0.3%	$\pm$0.6%	$\pm$3.0%	
	No-Tag	$\pm$1.9%	$\pm$5.7%	$\pm$3.8%	$\pm$1.9%	
Q^2	all		shape only	shape only		
ISR/FSR	all	shape only				
$b$ness Tag	HF-Tag	$\pm$7.8%	$\pm$7.8%	$\pm$9.2%	$\pm$7.6%	
	LF-Tag	$\pm$0.2%	$\pm$0.0%	$\pm$1.2%	$\pm$2.8%	
	No-Tag	$\pm$0.4%	$\pm$0.1%	$\pm$1.8%	$\pm$4.5%	
QG Tag	LF-Tag	$\pm$10%	$\pm$16%	$\pm$2.0%	$\pm$15%	
	No-Tag	$\pm$4.3%	$\pm$3.5%	$\pm$2.0%	$\pm$2.0%	
Lepton Energy Scale	all	$\pm$0.5%	$\pm$0.5%	$\pm$0.5%	$\pm$1.5%	
Lepton Energy Res.	all	$\pm$0.1%	$\pm$0.1%	$\pm$0.0%	$\pm$2.7%	

Table 8.3: Summary of the systematic uncertainties considered in the fit of the dijet mass distribution. Uncertainties that change both the shape and rate of templates used in the fit are treated in a correlated fashion.

8.2.6 Acceptance Uncertainties

Finally, there are additional systematic uncertainties not included in the fit itself but that affect our acceptance (and thus our measured cross section):

- Lepton Trigger/Reconstruction Efficiencies: We assign a 2.5% uncertainty on the trigger and reconstruction efficiencies we calculate by comparing data and MC in the Z+ 1 jet region, based on a collection of the statistical uncertainties for the lepton pair types.

- Parton Distribution Functions (PDFs): For our signal template, we take a 2% uncertainty on the normalization due to uncertainties in the parton distribution functions, as done in [31].

- Luminosity: We take a 6% uncertainty on the luminosity for all MC-based templates, as suggested by [77].

These uncertainties along with those summarized in Tab. 8.3 are used in the generation of pseudo-data, which are in turn used to estimate our sensitivity and get our final measurement.

8.3 Results

We perform the fit to the data in the three fitting channels as described above. The background + signal model, after the fit, is compared to the data in Fig. 8.7. The number of events in each channel after the fit is summarized in Tab. 8.4. We find that in the best fit to the data, we fit for $\approx$ 50% of the expected signal normalization. There is good agreement between data and MC in each of the three fitting channels, with a total $\chi^2_{\text{test}}/\text{d.o.f} = 59.8/55$, which has a corresponding p-value of 0.31. However, the fit assuming no signal has $\chi^2_{\text{null}}/\text{d.o.f} = 62.0/56$, and so is itself not a bad fit—the p-value for this no-signal fit is 0.27.

The changes to the template normalizations and the systematic parameters in the fit are summarized in Tab. 8.5. The errors are as reported by the minimization program `MINIUT` [78], which is used to perform the χ^2 minimization in `mclimit`, and should be

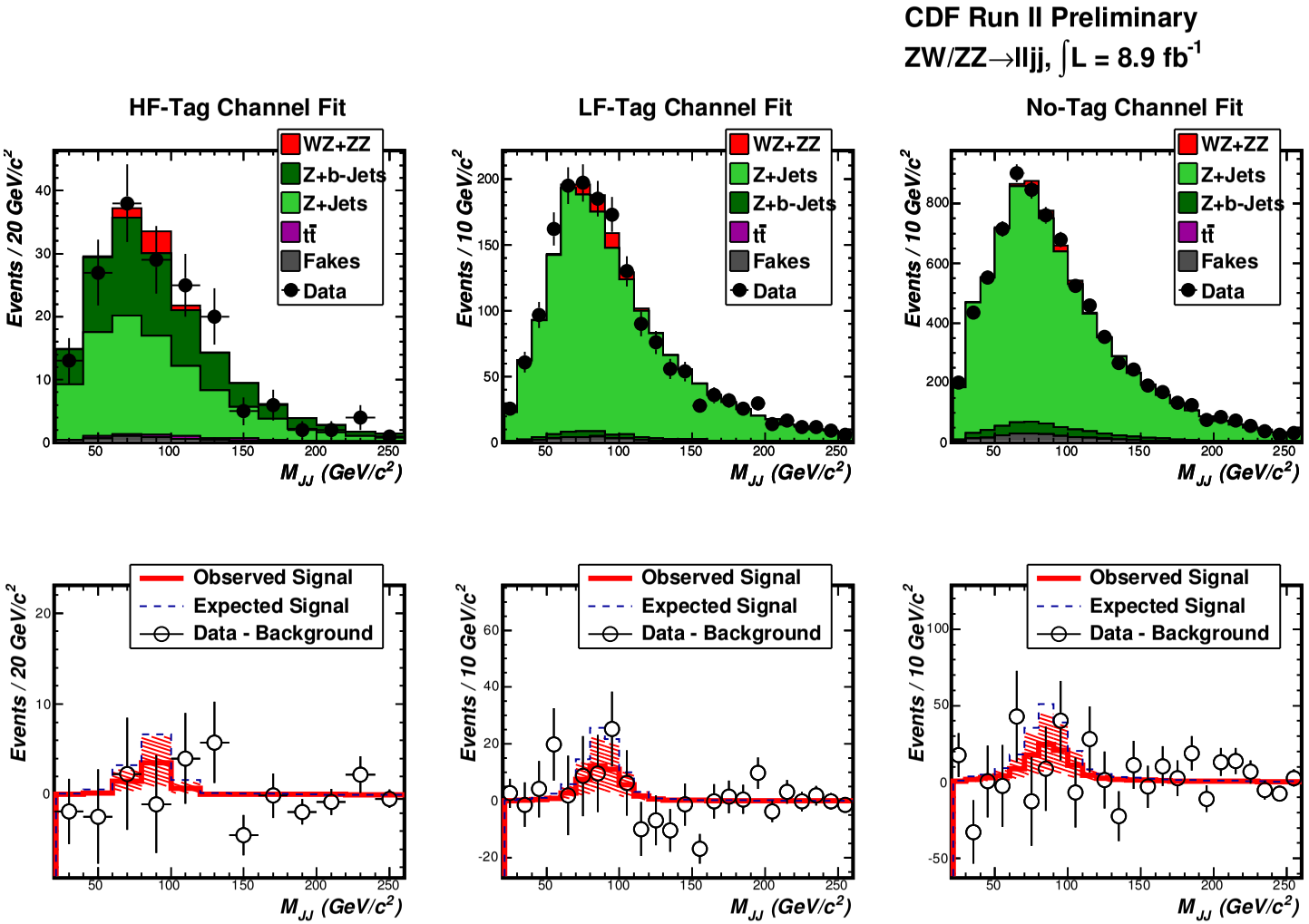

Figure 8.7: Result of the fit to data for ZW/ZZ in our dilepton + dijet selection in our heavy-flavor tagged channel (left), light-flavor tagged channel (center), and untagged channel (right). The top row shows the output from the fit compared to the data, while the bottom row shows the background subtracted from data, compared to the expected (red dashed line) and fitted (blue solid line, with uncertainties in blue bands) signal contributions.

Process	N_{events}, HF-Tag	N_{events}, LF-Tag	N_{events}, No-Tag
Z + jets	84 ± 11	1600 ± 59	7160 ± 660
Z + b jets	78 ± 33	42 ± 20	400 ± 170
$t\bar{t}$	3.22 ± 0.35	0.73 ± 0.07	5.22 ± 0.43
Fakes	4.6 ± 2.4	39 ± 20	273 ± 142
Total Bkg.	170 ± 26	1682 ± 38	7840 ± 530
$ZW + ZZ$	6.1 ± 4.7	44 ± 32	104 ± 77
Total Events	176 ± 26	1726 ± 42	7940 ± 530
Data Events	172	1724	7950

Table 8.4: The number of events in from each template, in each channel, after performing the fit to the data.

taken to be approximations. From the parameters of the fit, we can see that the diboson normalization is half of that expected from the Standard Model prediction. Additionally, the Z + b jets normalization is different (by almost 1σ) from its predicted value.

Template	Fit Normalization	$\pm$	Error
Z+ jets	+4.4%	$\pm$	4.6%
Z + b jets	-37%	$\pm$	19%
$t\bar{t}$	-0.0%	$\pm$	6.5%
Fakes	-4%	$\pm$	49%
ZW/ZZ	-50%	$\pm$	34%
Systematic	Fit Value (σ)	$\pm$	Error (σ)
Jet Resolution	+0.11	$\pm$	0.84
Jet Energy Scale	+0.35	$\pm$	0.72
Q^2	-1.31	$\pm$	0.46
$b$ness Tag/Mistag	+0.48	$\pm$	0.95
QG Tag/Mistag	-0.04	$\pm$	0.12
Lepton Energy Scale	+0.00	$\pm$	1.00
Lepton Energy Resolution	+0.00	$\pm$	1.00
ISR/FSR	-0.04	$\pm$	0.99

Table 8.5: The adjustments to the template normalizations and nuisance parameters from the fit. The errors are as reported by the minimization program `MINUIT`.

Outside of the template normalizations, however, we see some interesting behavior in

some of the nuisance parameters. The jet energy resolution and jet energy scale are both pulled from their nominal values, in a way that would broaden the dijet mass distribution of the Z+ jets background. Additionally, the Q^2 systematic parameter varies greatly, moving 1.3σ away from the central value in a way that broadens the Z+ jets spectrum. These shifts are largely driven by the dijet mass shape appearing broader in data than MC, seen in Figs. 4.7 and 8.4-8.6, especially in the untagged channel (Fig. 8.6). While the modeling of the jet energies is improved by the independent quark and gluon energy corrections (see Sec. 5.6), the improvement in the dijet mass shape is not nearly as dramatic. Still, despite some of the pulls from fit, the overall fit is good.

8.3.1 Cross-Checks and Additional Fits

We perform a number of checks to be sure that our fit behaves in a the way we expect. Among these is performing fits on a set of pseudo-data, generated using Poisson fluctuations on the number of events from the input templates with Gaussian variations of the nuisance parameters. We check the output fit parameters from these pseudo-experiments, and then check the difference between the thrown nuisance parameter (used to generate the pseudo-data) and the value extracted from the fit. If the mean of this pull differs significantly from 0, our fitter is biased in some direction. The width of the distribution should also be unity: if it is smaller than 1, the fitter is performing better than expected at getting the correct thrown value; if it is larger than 1, than our fitter is less sensitive than expected, and we thus may be underestimating that systematic error.

The fit values and pulls are shown in App. E. We see that the pull distributions are centered around zero, and the the widths are 1 or less, indicating that we are not underestimating any systematic uncertainties.

In addition, we perform a number of other fits to the data to ensure that our results are

internally consistent, and to perhaps shed some light on the behavior of the fit, and see what may be driving the result we obtain. These additional fits are shown in App. F. They include the following items of interest:

- In fits to the electron-only data and muon-only data, we find that the electron-only data fit (Sec. F.1) returns a diboson signal normalization rather close to the expected one, down only by 7%. In the muon-only data fit (Sec. F.2), we fit for no signal. However, the error on these normalizations is fairly large ($\pm$46% in the electron fit, and $\pm$42% in the muon fit, as reported by MINUIT), and so they are broadly consistent with each other. We do find that the muon-only fit is being dominated by behavior in the untagged channel, as when we remove that channel we obtain a normalization nearer to the Standard Model expectation ($-13\% \pm 38\%$, see Sec. F.3). Importantly, we do not see large, inconsistent shifts in the nuisance parameters when fitting for electron and muon data separately.

- We also perform fits in each of the tag categories independently, in Secs. F.4-F.6. We find that the number of events in the heavy-flavor tag fit is low, thus giving large uncertainties on the signal normalization fit. The fit in the light-flavor tag region has a signal normalization 26% less than the expected cross section, but overall is rather good and the nuisance parameters come in very near their predicted values, with the exception of the Q^2 parameter affecting our ALPGEN samples. The untagged channel fit sees the largest differences in the background model from the nominal templates, where both the jet energy scale and Q^2 systematics move more dramatically to broaden the m_{jj} distribution in MC. This is consistent with our previous findings that quark jets are typically well-modeled in MC, while there are problems with the modeling of gluon jets, which are far more prominent in the untagged channel.

- Finally, we perform a fit where we constrain the normalization of the ZW/ZZ template to be within 6% of the expected value. This checks for consistency of our model with the SM, and allows us to see if the nuisance parameter shifts in Tab. 8.5 are a result of the low signal normalization. We find that the χ^2/d.o.f for the fit with constrained signal is 61.8/56, similar to the quality of the null hypothesis fit. The fits to the nuisance parameters come out similar to the fit with the signal normalization unconstrained, suggesting that nuisance parameter fit values like the Q^2 and JES are largely being driven by changes to improve agreement to the background model in regions of low and/or high dijet mass.

8.3.2 Cross Section Measurement

To translate the result of our fit to the data to bounds or limits on the cross section of ZW/ZZ production, we construct Feldman-Cousins bands [79] by analyzing the distribution of fitted (*i.e.*, measured) cross sections in pseudo-experiments generated with a variety of scale factors on the input signal cross section, $k = \sigma_{\text{generated}}/\sigma_{SM}$. The set of input cross sections in our pseudo-experiments range from 0.0 to 2.9 $\times \sigma_{SM}$ with a step size of 0.1 $\times \sigma_{SM}$.

For each set of pseudo-experiments, we find a range of measured cross sections that meets a desired coverage threshold. To do this, we first bin the measured scale factors on σ_{SM} in a histogram. The bin containing the input value of the σ_{SM} scale factor acts as a seed for our coverage interval. We then check the bin contents directly above and below the current interval, and add the most populous one (the one with the highest probability content) to form a new coverage interval, repeating this process until the desired coverage is achieved. Because the fitter cannot return a negative number of fitted signal events, the first bin (at a measured scale factor of 0) may contain a very large number of events;

however, we do not treat it differently than any other bin, and if it is included in the range, its entire contents contribute towards the calculation of coverage. This method is a slight variation of the method proposed in [79], due to the strict boundary on the measured scale factors (as well as the input scale factors), but retains the properties that it avoids flip-flopping, and aims for coverage as close as possible to (but always as much as) the stated value.

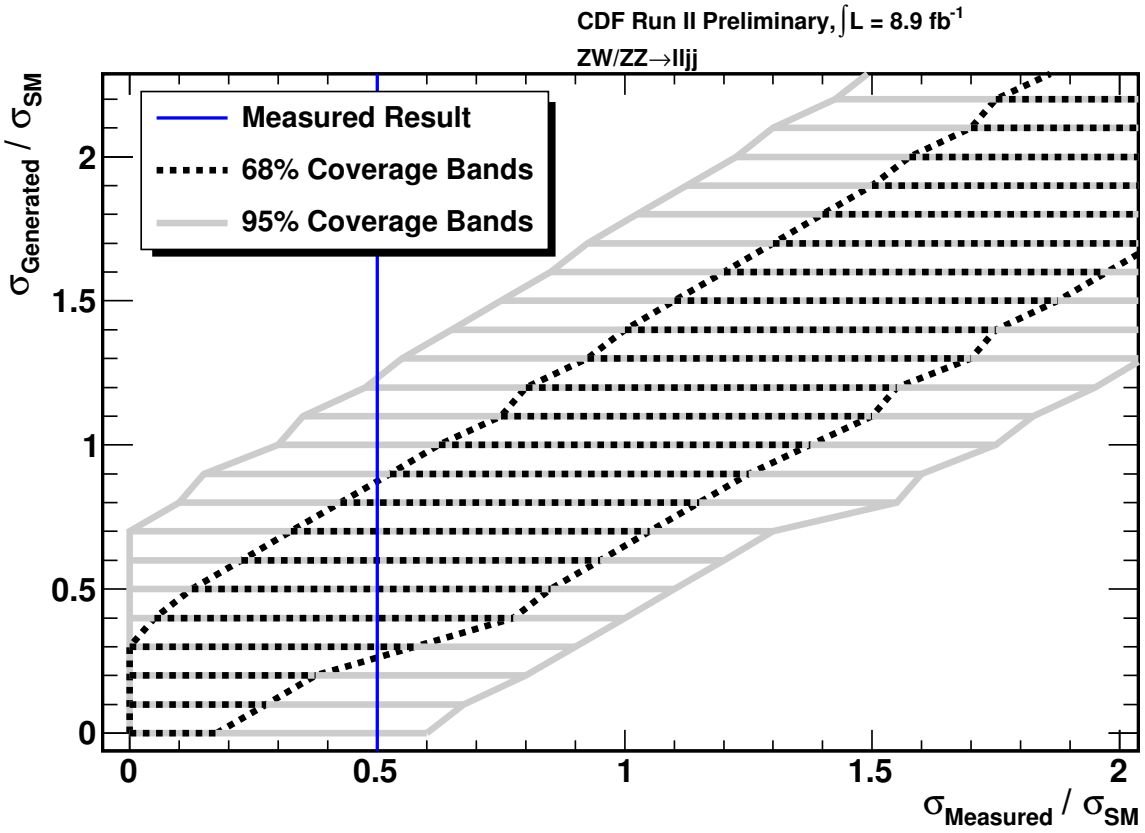

Figure 8.8: Condence bands showing the expected range of measured cross sections as a function of the true cross section, with 68% CL (black dashed region) and 95% CL (solid gray region). Our measured result of $\sigma(p\bar{p} \to ZW/ZZ) = 2.5^{+2.0}_{-1.0}$ pb corresponds to a limit of $\sigma_{ZW/ZZ} < 6.1$ pb ($1.2 \times \sigma_{SM}$) at the 95% C.L.

Fig. 8.8 shows the results of our Feldman-Cousins analysis. On the *x*-axis is the measured cross section, as determined from the fit; on the *y*-axis is the signal cross section in the generated pseudo-data. Each band on the *y*-axis represents the measured

cross sections obtained in a set of 50,000 pseudo-experiments with a given input signal cross section, with either 68% coverage (1σ) or 95% coverage (2σ). We may then read off 1σ bounds on the measured cross section by drawing a vertical line at the desired cross section and noting where that line intersects with the 68% coverage bands. Similarly, we may establish 2σ bounds by noting intersection with the 95% coverage bands. If the 95% bands extend down to $\sigma_{\text{generated}}/\sigma_{\text{SM}} = 0.0$, then our measured result is consistent with no observed signal, and we then may only report an upper limit at the 95% confidence level (CL).

A vertical line is drawn in Fig. 8.8 that indicates our measured result. Using the 1σ bands, we measure $\sigma(p\bar{p} \rightarrow ZW/ZZ) = 2.5^{+2.0}_{-1.0}$ pb, compared to the standard model prediction of $\sigma_{SM} = 5.1$ pb. We do not exclude the no-signal hypothesis at the 95% CL—we do so at 92% CL, corresponding to a 1.75σ significance—and establish a limit of $\sigma_{ZW/ZZ} < 6.1$ pb ($1.2 \times \sigma_{SM}$). The expected measurement, based on drawing a vertical line at $\sigma_{\text{measured}}/\sigma_{\text{SM}} = 1.0$, is $\sigma_{ZW/ZZ,\text{exp.}} = 5.1 \pm 2.0$ pb.

CHAPTER 9

CONCLUSION

In this thesis, we describe a search for ZW/ZZ diboson production in a final state with two charged high-p_T leptons and two hadronic jets at the CDF detector. These processes, predicted by the Standard Model are understood well and experimentally confirmed to exist in fully-leptonic final states, and therefore offer a chance to test our models of some of the basic interactions of particle physics. Furthermore, the close relationship of this final state to Higgs boson production (in association with a Z boson) and $H \to b\bar{b}$ decay make this analysis even more important to conduct—not only to prove we can find two jets from the decay of a massive boson, but also to prove that we can understand the background processes in this channel.

In this search, we have done a number of new things to both improve the modeling of our simulations and increase the acceptance to our diboson signal. We have increased the number of triggers from which we get the data and loosened many of the quality requirements on our lepton identification in order to increase acceptance, making use of the fact that $Z \to \ell^+\ell^-$ decays are well-understood and may be used to calibrate the efficiency at which we accept events with these decays. We further use leptonic Z decays to calculate new corrections encapsulating the difference in the jet energy scale for quark and gluon jets in our MC simulations. We use neural network discriminants to tag jets likely coming from heavy and light-flavor quarks, and ensure that those discriminants our well-modeled by our simulations. Both taggers use the large number of top quark events now available in the CDF data sample as a way to check our simulations, and these samples may be used to derive tag efficiencies and mistag rates. We use yesterday's

discoveries—like the W and Z bosons, and the top quark—as a benchmark for testing the new tools that serve as the drivers for our new searches.

Using data over the entirety of CDF Run II, an integrated luminosity of 8.9 fb^{-1}, we fit for the normalization of $ZW/ZZ \to \ell^+\ell^-$ + $q\bar{q}'$ events using the dijet invariant mass distribution. While we expected a measurement of σ_{ZW+ZZ} = 5.1 $\pm$ 2.0 pb, corresponding to a significance of about 2.6σ, we measured a cross section of

$$\sigma_{ZW+ZZ} = 2.5^{+2.0}_{-1.0} \text{ pb.}$$

This result is significant at about 1.75σ. We also report a limit on the cross section:

$$\sigma_{ZW+ZZ} < 6.1 \text{ pb at 95\% CL.}$$

This measurement is consistent with the Standard Model prediction.

Combining this result with other decay channels, like $W^{\pm}Z \to \ell^{\pm}\nu_{\ell}$ + $q\bar{q}$ and $ZZ \to \nu\bar{\nu}$ + $q\bar{q}$, will lead to a more significant measurement of the WZ and ZZ diboson cross sections. If searches in these three main channels (charged leptons plus jets, one lepton plus $\not\!\!E_T$ plus jets, and $\not\!\!E_T$ plus jets with no charged leptons) have a similar expected sensitivity of about 2.5σ, then the overall expected sensitivity would be about 4.3σ assuming independent treatment of uncertainties. In truth, many uncertainties are correlated, which may both hurt and help the expected sensitivity. While combinations of final states with b jets have been done (see [35, 36]), the combination over all semi-hadronic WZ/ZZ decays has not yet been reported. Future work based on data from the Tevatron and the LHC will lead to discoveries of WZ and ZZ production in these particular decay channels alongside a much deeper understand of the W/Z + jets backgrounds, which may be essential to searches for physics both in and beyond the Standard Model in similar final states.

APPENDIX A

MC SAMPLES

Here we list the MC samples used at various points in this analysis. k-factors, which account for corrections to the leading-order cross section calculations, are included in the cross section whenever necessary (for all samples except the diboson and $t\bar{t}$ samples). High luminosity samples with a higher N_{vtx} distribution are used when available. The ALPGEN samples labelled "Drell-Yan" in this table refer to either a low mass mediator (20 $GeV/c^2 < m_{\gamma/Z^*} < 75$ GeV/c^2, the xt* samples), or one with high mass (105 $GeV/c^2 < m_{\gamma/Z^*} < 600$ GeV/c^2, the yt* samples). The ALPGEN Z samples have a m_Z window between 75 and 105 GeV/c^2.

CDF Catalog Name	σ (pb)	MC Generator	Process
it0szz (ht0szz)	1.38	PYTHIA (High Lumi)	ZZ
it0swz (ht0swz)	3.70	PYTHIA (High Lumi)	WZ
it0sww (ht0sww)	12.4	PYTHIA (High Lumi)	WW
tt1s25	7.5	PYTHIA	$t\bar{t}$
zt0sp0 (bt0sz0)	221.2	ALPGEN (High Lumi)	$Z \rightarrow e^+e^-$ (+0p)

Continued on next page ...

CDF Catalog Name	σ (pb)	MC Generator	Process
zt0sp1 (bt0sz1)	30.24	ALPGEN (High Lumi)	$Z \rightarrow e^+e^-$ (+1p)
zt0szb (bt0sz2)	4.844	ALPGEN (High Lumi)	$Z \rightarrow e^+e^-$ (+2p)
zt0s3p (bt0sz3)	0.77	ALPGEN (High Lumi)	$Z \rightarrow e^+e^-$ (+3p)
zt0s4p (bt0sz4)	0.1389	ALPGEN (High Lumi)	$Z \rightarrow e^+e^-$ (+$\geq$4p)
zt0sp5 (bt0sz5)	221.2	ALPGEN (High Lumi)	$Z \rightarrow \mu^+\mu^-$ (+0p)
zt0sp6 (bt0sz6)	30.24	ALPGEN (High Lumi)	$Z \rightarrow \mu^+\mu^-$ (+1p)
zt0szt (bt0sz7)	4.844	ALPGEN (High Lumi)	$Z \rightarrow \mu^+\mu^-$ (+2p)
zt0s8p (bt0sz8)	0.7672	ALPGEN (High Lumi)	$Z \rightarrow \mu^+\mu^-$ (+3p)
zt0s9p (bt0sz9)	0.1389	ALPGEN (High Lumi)	$Z \rightarrow \mu^+\mu^-$ (+$\geq$4p)
zt0st2 (bt0sza)	221.2	ALPGEN (High Lumi)	$Z \rightarrow \tau^+\tau^-$ (+0p)
zt0st3 (bt0szb)	30.24	ALPGEN (High Lumi)	$Z \rightarrow \tau^+\tau^-$ (+1p)
zt0st4 (bt0szc)	5.796	ALPGEN (High Lumi)	$Z \rightarrow \tau^+\tau^-$ (+$\geq$2p)
zt0sb0 (bt0szd)	1.403	ALPGEN (High Lumi)	$Z \rightarrow e^+e^- + b\bar{b}$ (+0p)
zt0sb1 (bt0sze)	0.3752	ALPGEN (High Lumi)	$Z \rightarrow e^+e^- + b\bar{b}$ (+1p)
zt0sb2 (bt0szf)	0.1078	ALPGEN (High Lumi)	$Z \rightarrow e^+e^- + b\bar{b}$ (+$\geq$2p)
zt0sb5 (bt0szg)	1.403	ALPGEN (High Lumi)	$Z \rightarrow \mu^+\mu^- + b\bar{b}$ (+0p)
zt0sb6 (bt0szh)	0.3752	ALPGEN (High Lumi)	$Z \rightarrow \mu^+\mu^- + b\bar{b}$ (+1p)
zt0sb7 (bt0szi)	0.1078	ALPGEN (High Lumi)	$Z \rightarrow \mu^+\mu^- + b\bar{b}$ (+$\geq$2p)
zt0sbt (bt0szj)	1.75	ALPGEN (High Lumi)	$Z \rightarrow \tau^+\tau^- + b\bar{b}$ (+$\geq$0p)

Continued on next page . . .

CDF Catalog Name	σ (pb)	MC Generator	Process
xt0s0p (zt0so6)	224	ALPGEN (High Lumi)	Drell-Yan$\rightarrow e^+e^-$ (+0p)
xt0s1p (zt0so7)	11.746	ALPGEN (High Lumi)	Drell-Yan$\rightarrow e^+e^-$ (+1p)
xt0s2p (zt0so9)	2.254	ALPGEN (High Lumi)	Drell-Yan$\rightarrow e^+e^-$ (+2p)
xt0s3p (zt0soa)	0.3262	ALPGEN (High Lumi)	Drell-Yan$\rightarrow e^+e^-$ (+3p)
xt0s4p (zt0sob)	0.5572	ALPGEN (High Lumi)	Drell-Yan$\rightarrow e^+e^-$ (+$\geq$4p)
xt0s5p (zt0soc)	224	ALPGEN (High Lumi)	Drell-Yan$\rightarrow \mu^+\mu^-$ (+0p)
xt0s6p (zt0sod)	11.746	ALPGEN (High Lumi)	Drell-Yan$\rightarrow \mu^+\mu^-$ (+1p)
xt0s7p (zt0sof)	2.254	ALPGEN (High Lumi)	Drell-Yan$\rightarrow \mu^+\mu^-$ (+2p)
xt0s8p (zt0sog)	0.3262	ALPGEN (High Lumi)	Drell-Yan$\rightarrow \mu^+\mu^-$ (+3p)
xt0s9p (zt0soh)	0.5572	ALPGEN (High Lumi)	Drell-Yan$\rightarrow \mu^+\mu^-$ (+$\geq$4p)
xt0st0 (zt0soi)	224	ALPGEN (High Lumi)	Drell-Yan$\rightarrow \tau^+\tau^-$ (+0p)
xt0st1 (zt0soj)	11.732	ALPGEN (High Lumi)	Drell-Yan$\rightarrow \tau^+\tau^-$ (+1p)
xt0st2 (zt0sok)	2.548	ALPGEN (High Lumi)	Drell-Yan$\rightarrow \tau^+\tau^-$ (+$\geq$2p)
yt0s0p (zt0sol)	5.698	ALPGEN (High Lumi)	Drell-Yan$\rightarrow e^+e^-$ (+0p)
yt0s1p (zt0som)	0.9884	ALPGEN (High Lumi)	Drell-Yan$\rightarrow e^+e^-$ (+1p)
yt0s2p (zt0son)	0.1638	ALPGEN (High Lumi)	Drell-Yan$\rightarrow e^+e^-$ (+2p)
yt0s3p	0.0259	ALPGEN	Drell-Yan$\rightarrow e^+e^-$ (+3p)
zt0sop	4.66×10^{-3}	ALPGEN, High Lumi	Drell-Yan$\rightarrow e^+e^-$ (+$\geq$4p)
yt0s5p (zt0soq)	5.698	ALPGEN (High Lumi)	Drell-Yan$\rightarrow \mu^+\mu^-$ (+0p)

Continued on next page . . .

CDF Catalog Name	σ (pb)	MC Generator	Process
yt0s6p (zt0sor)	0.9884	ALPGEN (High Lumi)	Drell-Yan$\rightarrow\mu^+\mu^-$ (+1p)
yt0s7p (zt0sos)	0.1638	ALPGEN (High Lumi)	Drell-Yan$\rightarrow\mu^+\mu^-$ (+2p)
yt0s8p (zt0sot)	0.0259	ALPGEN (High Lumi)	Drell-Yan$\rightarrow\mu^+\mu^-$ (+3p)
zt0sou	4.65×10^{-3}	ALPGEN, High Lumi	Drell-Yan$\rightarrow\mu^+\mu^-$ (+$\geq$4p)
zt0s0h (zt0sov)	5.698	ALPGEN (High Lumi)	Drell-Yan$\rightarrow\tau^+\tau^-$ (+0p)
zt0s1h (zt0sow)	0.9898	ALPGEN (High Lumi)	Drell-Yan$\rightarrow\tau^+\tau^-$ (+1p)
zt0s2h (zt0sox)	0.1638	ALPGEN (High Lumi)	Drell-Yan$\rightarrow\tau^+\tau^-$ (+2p)
zt0s3h (zt0soy)	0.0259	ALPGEN (High Lumi)	Drell-Yan$\rightarrow\tau^+\tau^-$ (+3p)
zt0s4h (zt0soz)	4.62×10^{-3}	ALPGEN (High Lumi)	Drell-Yan$\rightarrow\tau^+\tau^-$ (+$\geq$4p)
xt0sb0	0.8204	ALPGEN	Drell-Yan$\rightarrow e^+e^-+b\bar{b}$ (+0p)
xt0sb1	0.1638	ALPGEN	Drell-Yan$\rightarrow e^+e^-+b\bar{b}$ (+1p)
xt0sb2	0.0442	ALPGEN	Drell-Yan$\rightarrow e^+e^-+b\bar{b}$ (+$\geq$2p)
xt0sb5	0.8204	ALPGEN	Drell-Yan$\rightarrow\mu^+\mu^-+b\bar{b}$ (+0p)
xt0sb6	0.1638	ALPGEN	Drell-Yan$\rightarrow\mu^+\mu^-+b\bar{b}$ (+1p)
xt0sb7	0.0442	ALPGEN	Drell-Yan$\rightarrow\mu^+\mu^-+b\bar{b}$ (+$\geq$2p)
xt0sbt	0.8764	ALPGEN	Drell-Yan$\rightarrow\tau^+\tau^-+b\bar{b}$ (+$\geq$0p)
yt0s0b	0.0403	ALPGEN	Drell-Yan$\rightarrow e^+e^-+b\bar{b}$ (+0p)
yt0s1b	0.0118	ALPGEN	Drell-Yan$\rightarrow e^+e^-+b\bar{b}$ (+1p)
yt0s2b	3.36×10^{-3}	ALPGEN	Drell-Yan$\rightarrow e^+e^-+b\bar{b}$ (+$\geq$2p)

Continued on next page ...

Table A.1 – Continued

CDF Catalog Name	σ (pb)	MC Generator	Process
yt0s5b	0.0403	ALPGEN	Drell-Yan$\rightarrow\mu^+\mu^-$+$b\bar{b}$ (+0p)
yt0s6b	0.0118	ALPGEN	Drell-Yan$\rightarrow\mu^+\mu^-$+$b\bar{b}$ (+1p)
yt0s7b	3.36×10^{-3}	ALPGEN	Drell-Yan$\rightarrow\mu^+\mu^-$+$b\bar{b}$ (+$\geq$2p)
yt0stb	0.02534	ALPGEN	Drell-Yan$\rightarrow\tau^+\tau^-$+$b\bar{b}$ (+$\geq$0p)
pt0sw0 (ut0s00)	2520	ALPGEN (High Lumi)	$W^\pm \rightarrow e^\pm\nu_e$ (+0p)
pt0sw1 (ut0s01)	315	ALPGEN (High Lumi)	$W^\pm \rightarrow e^\pm\nu_e$ (+1p)
pt0s2w (ut0s02)	49.42	ALPGEN (High Lumi)	$W^\pm \rightarrow e^\pm\nu_e$ (+2p)
pt0s3w (ut0s03)	7.826	ALPGEN (High Lumi)	$W^\pm \rightarrow e^\pm\nu_e$ (+3p)
pt0s4w (ut0s04)	1.442	ALPGEN (High Lumi)	$W^\pm \rightarrow e^\pm\nu_e$ (+$\geq$4p)
pt0sw5 (ut0s05)	2520	ALPGEN (High Lumi)	$W^\pm \rightarrow \mu^\pm\nu_\mu$ (+0p)
pt0sw6 (ut0s06)	315	ALPGEN (High Lumi)	$W^\pm \rightarrow \mu^\pm\nu_\mu$ (+1p)
pt0s7w (ut0s07)	49.42	ALPGEN (High Lumi)	$W^\pm \rightarrow \mu^\pm\nu_\mu$ (+2p)
pt0s8w (ut0s08)	7.826	ALPGEN (High Lumi)	$W^\pm \rightarrow \mu^\pm\nu_\mu$ (+3p)
pt0s9w (ut0s09)	1.442	ALPGEN (High Lumi)	$W^\pm \rightarrow \mu^\pm\nu_\mu$ (+$\geq$4p)
ut0sw0 (ut0s10)	2520	ALPGEN (High Lumi)	$W^\pm \rightarrow \tau^\pm\nu_\tau$ (+0p)
ut0sw1 (ut0s11)	315	ALPGEN (High Lumi)	$W^\pm \rightarrow \tau^\pm\nu_\tau$ (+1p)
ut0s2w (ut0s12)	49.42	ALPGEN (High Lumi)	$W^\pm \rightarrow \tau^\pm\nu_\tau$ (+2p)

Continued on next page . . .

Table A.1 – Continued

CDF Catalog Name	σ (pb)	MC Generator	Process
ut0s3w (ut0s13)	7.826	ALPGEN (High Lumi)	$W^{\pm} \to \tau^{\pm}\nu_{\tau}$ (+3p)
ut0s4w (ut0s14)	1.442	ALPGEN (High Lumi)	$W^{\pm} \to \tau^{\pm}\nu_{\tau}$ (+$\geq$4p)
bt0s0w (bt0s00)	5.841	ALPGEN (High Lumi)	$W^{\pm} \to e^{\pm}\nu_{e}$+$b\bar{b}$ (+0p)
bt0s1w (bt0s01)	1.740	ALPGEN (High Lumi)	$W^{\pm} \to e^{\pm}\nu_{e}$+$b\bar{b}$ (+1p)
bt0s2w (bt0s02)	0.5625	ALPGEN (High Lumi)	$W^{\pm} \to e^{\pm}\nu_{e}$+$b\bar{b}$ (+$\geq$2p)
bt0s5w (bt0s05)	5.841	ALPGEN (High Lumi)	$W^{\pm} \to \mu^{\pm}\nu_{\mu}$+$b\bar{b}$ (+0p)
bt0s6w (bt0s06)	1.740	ALPGEN (High Lumi)	$W^{\pm} \to \mu^{\pm}\nu_{\mu}$+$b\bar{b}$ (+1p)
bt0s7w (bt0s07)	0.5625	ALPGEN (High Lumi)	$W^{\pm} \to \mu^{\pm}\nu_{\mu}$+$b\bar{b}$ (+$\geq$2p)
dt0s0w (bt0s10)	5.841	ALPGEN (High Lumi)	$W^{\pm} \to \tau^{\pm}\nu_{\tau}$+$b\bar{b}$ (+0p)
dt0s1w (bt0s11)	1.740	ALPGEN (High Lumi)	$W^{\pm} \to \tau^{\pm}\nu_{\tau}$+$b\bar{b}$ (+1p)
dt0s2w (bt0s12)	0.5625	ALPGEN (High Lumi)	$W^{\pm} \to \tau^{\pm}\nu_{\tau}$+$b\bar{b}$ (+$\geq$2p)
zt0sq5	215.04	ALPGEN ($Q^2 \times 2.0$)	$Z \to \mu^{+}\mu^{-}$ (+0p)
zt0sq6	24.64	ALPGEN ($Q^2 \times 2.0$)	$Z \to \mu^{+}\mu^{-}$ (+1p)
zt0s7q	3.402	ALPGEN ($Q^2 \times 2.0$)	$Z \to \mu^{+}\mu^{-}$ (+2p)
zt0s8q	0.4634	ALPGEN ($Q^2 \times 2.0$)	$Z \to \mu^{+}\mu^{-}$ (+3p)
zt0s9q	0.0742	ALPGEN ($Q^2 \times 2.0$)	$Z \to \mu^{+}\mu^{-}$ (+$\geq$4p)
zt0sr5	237.86	ALPGEN ($Q^2 \times 0.5$)	$Z \to \mu^{+}\mu^{-}$ (+0p)

Continued on next page . . .

Table A.1 – Continued

CDF Catalog Name	σ (pb)	MC Generator	Process
zt0sr6	40.46	ALPGEN ($Q^2 \times 0.5$)	$Z \rightarrow \mu^+\mu^-$ (+1p)
zt0s7r	7.910	ALPGEN ($Q^2 \times 0.5$)	$Z \rightarrow \mu^+\mu^-$ (+2p)
zt0s8r	1.4742	ALPGEN ($Q^2 \times 0.5$)	$Z \rightarrow \mu^+\mu^-$ (+3p)
zt0s9r	0.301	ALPGEN ($Q^2 \times 0.5$)	$Z \rightarrow \mu^+\mu^-$ (+$\geq$4p)

Table A.1: List of Monte Carlo samples used in modeling our signals and backgrounds

APPENDIX B

KINEMATICS IN INDIVIDUAL FITTING CHANNELS

The figures in this appendix show the distributions in data and MC of various quantities in each of the individual fitting channels: the heavy-flavor tagged, light-flavor tagged, and untagged channels. Both distributions with the expectation from MC absolutely normalized and normalized to the data are shown. In the fit to the dijet mass distribution, the normalization of the Z+ jets sample is unconstrained in the fit, thus the agreement going into the fit will be more like the MC normalized to data.

B.1 Heavy-Flavor Tagged Events

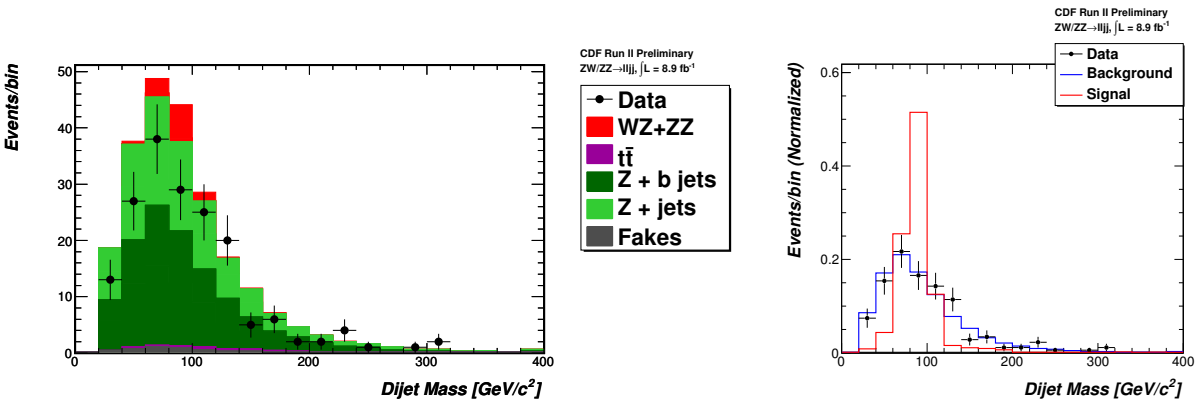

Figure B.1: The distribution of dijet mass in the Z + 2 jet signal region, with MC scaled to the data's luminosity (left) and with the distributions normalized to unity (right).

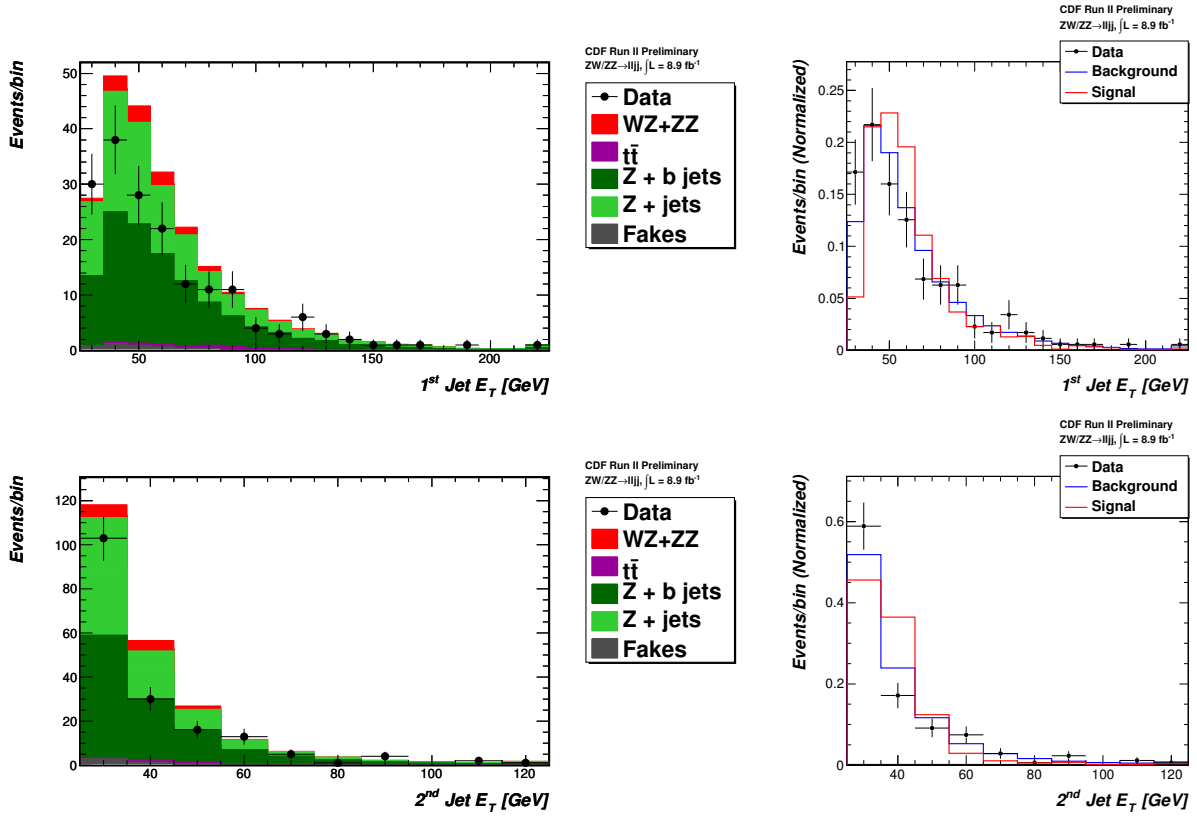

Figure B.2: The distribution of the leading jet E_T in the Z + 2 jet signal region, with MC scaled to the data's luminosity (top left) and with the distributions normalized to unity (top right), and the second leading jet E_T (bottom row).

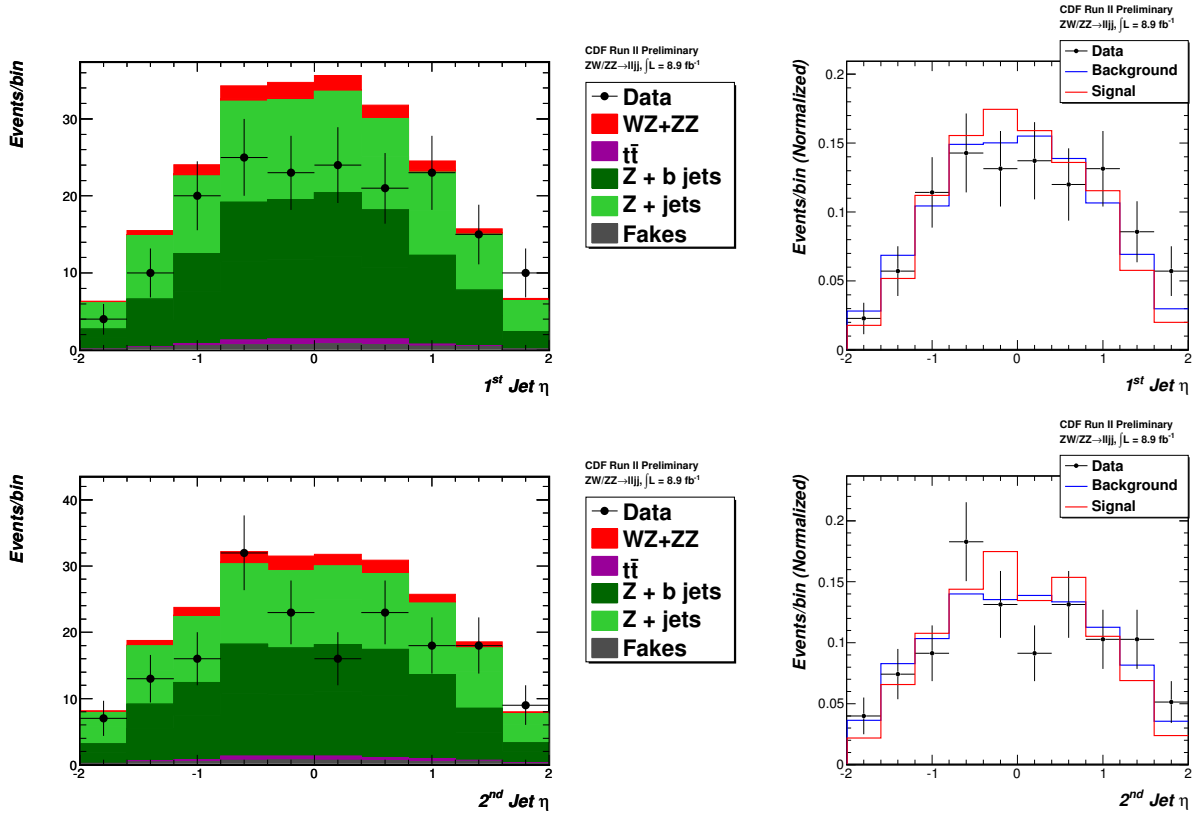

Figure B.3: The distribution of the leading jet η in the Z + 2 jet signal region, with MC scaled to the data's luminosity (top left) and with the distributions normalized to unity (top right), and the second leading jet η (bottom row).

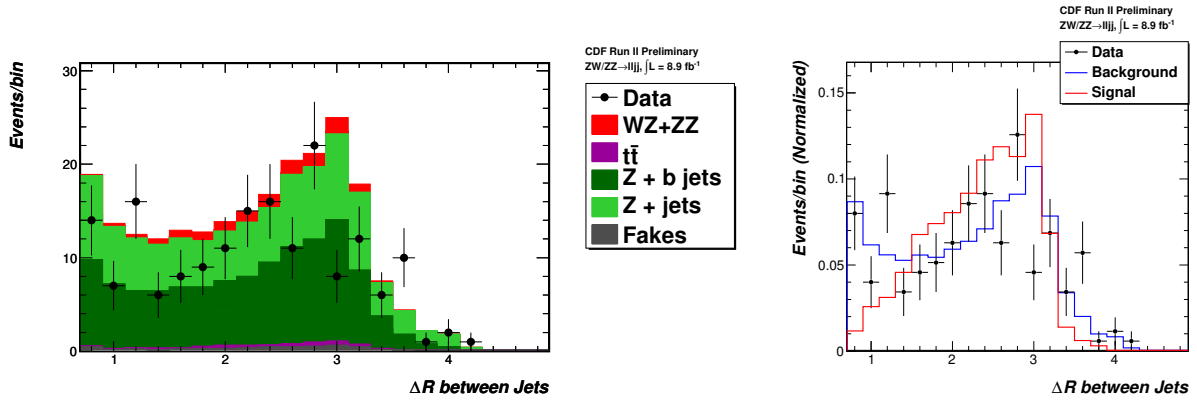

Figure B.4: The distribution of ΔR between jets in the Z + 2 jet signal region, with MC scaled to the data's luminosity (left) and with the distributions normalized to unity (right).

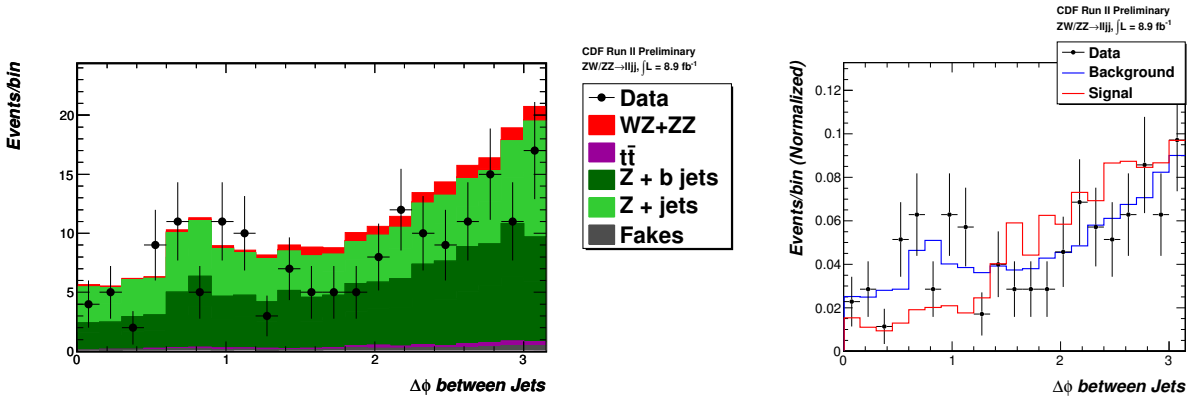

Figure B.5: The distribution of $\Delta\phi$ between jets in the Z + 2 jet signal region, with MC scaled to the data's luminosity (left) and with the distributions normalized to unity (right).

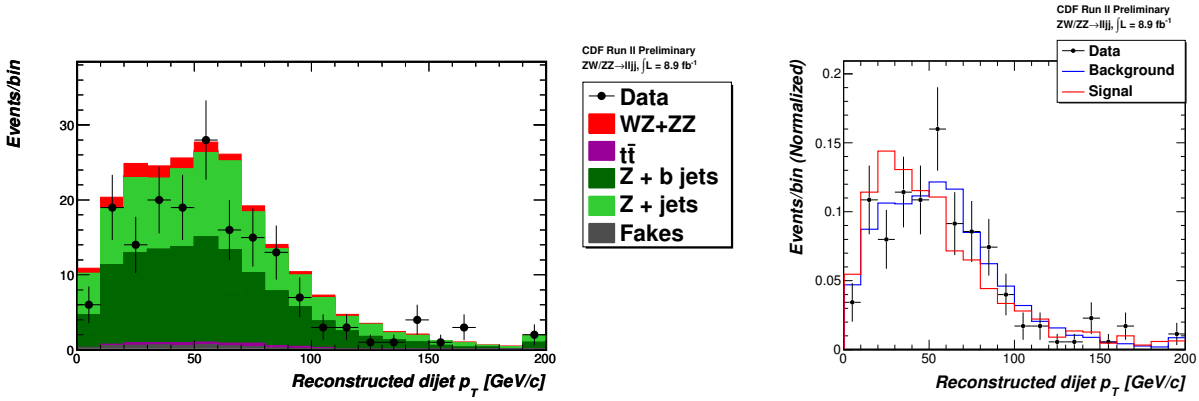

Figure B.6: The distribution of dijet p_T in the Z + 2 jet signal region, with MC scaled to the data's luminosity (left) and with the distributions normalized to unity (right).

B.2 Light-Flavor Tagged Events

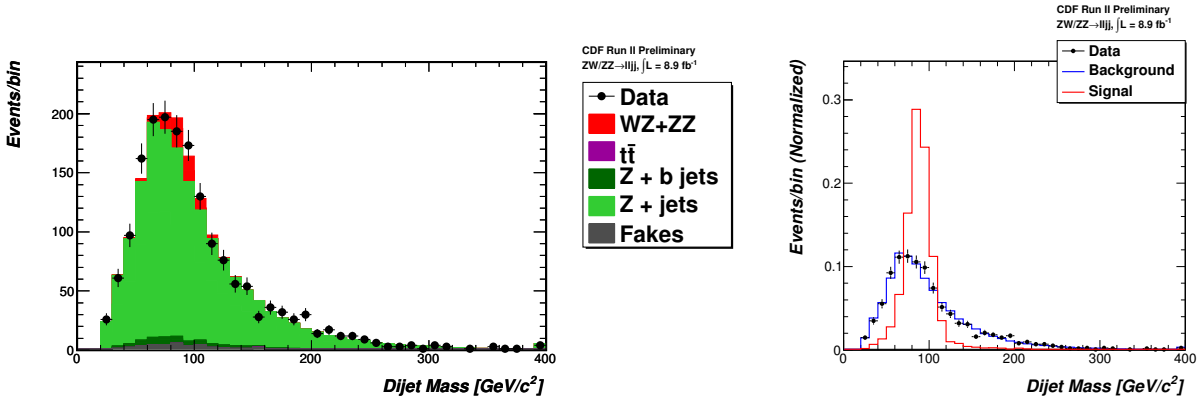

Figure B.7: The distribution of dijet mass in the Z + 2 jet signal region, with MC scaled to the data's luminosity (left) and with the distributions normalized to unity (right).

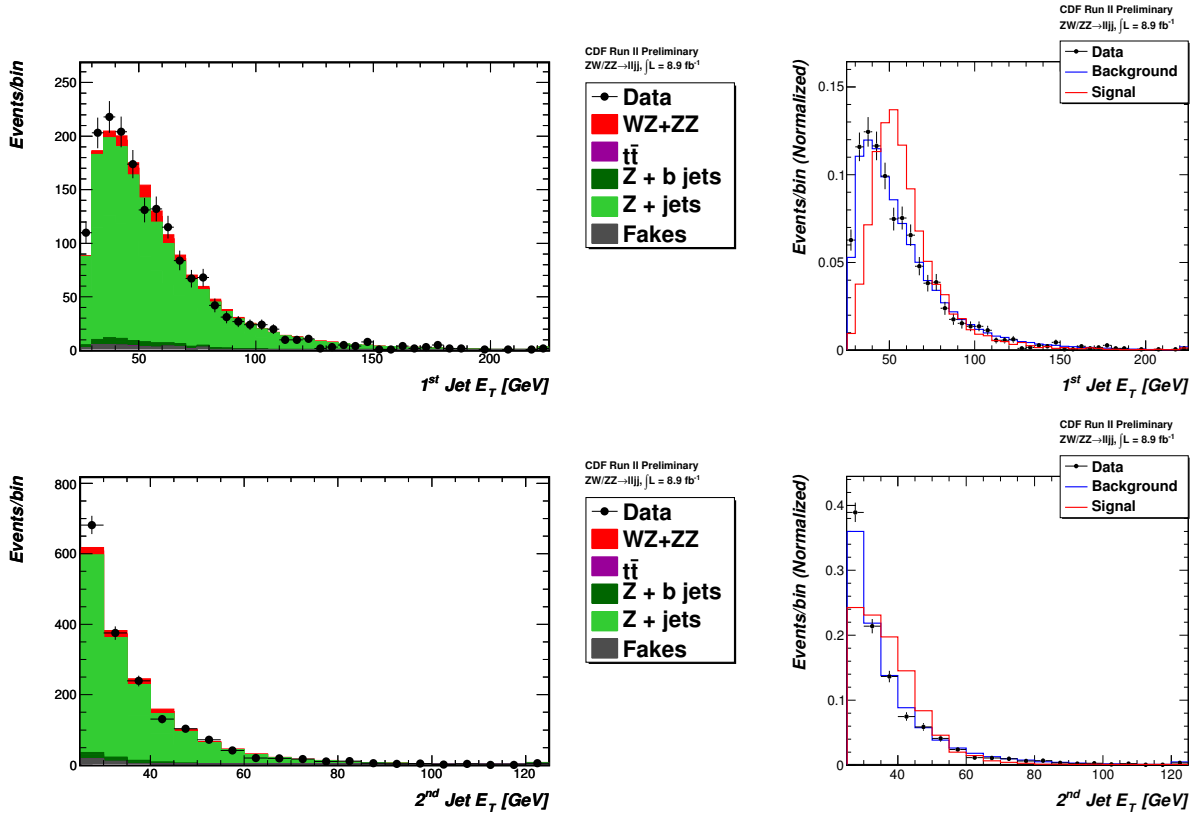

Figure B.8: The distribution of the leading jet E_T in the Z + 2 jet signal region, with MC scaled to the data's luminosity (top left) and with the distributions normalized to unity (top right), and the second leading jet E_T (bottom row).

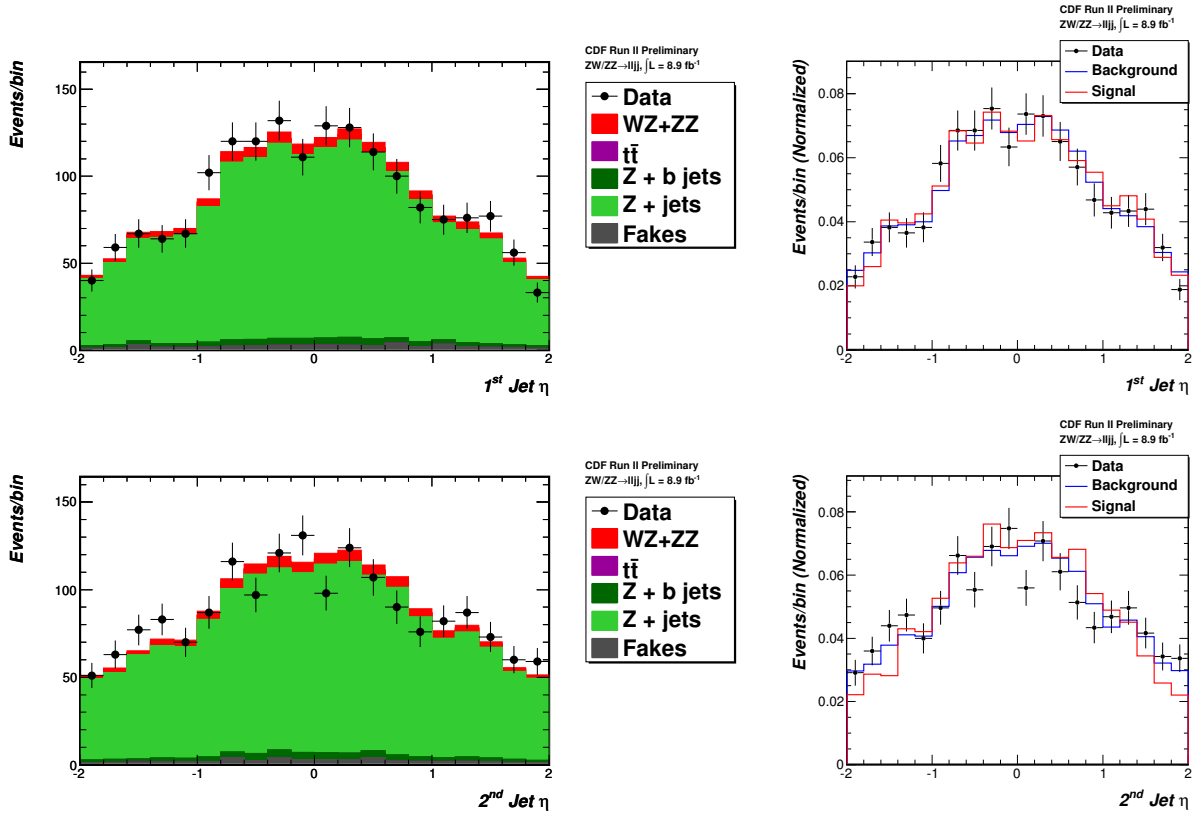

Figure B.9: The distribution of the leading jet η in the Z + 2 jet signal region, with MC scaled to the data's luminosity (top left) and with the distributions normalized to unity (top right), and the second leading jet η (bottom row).

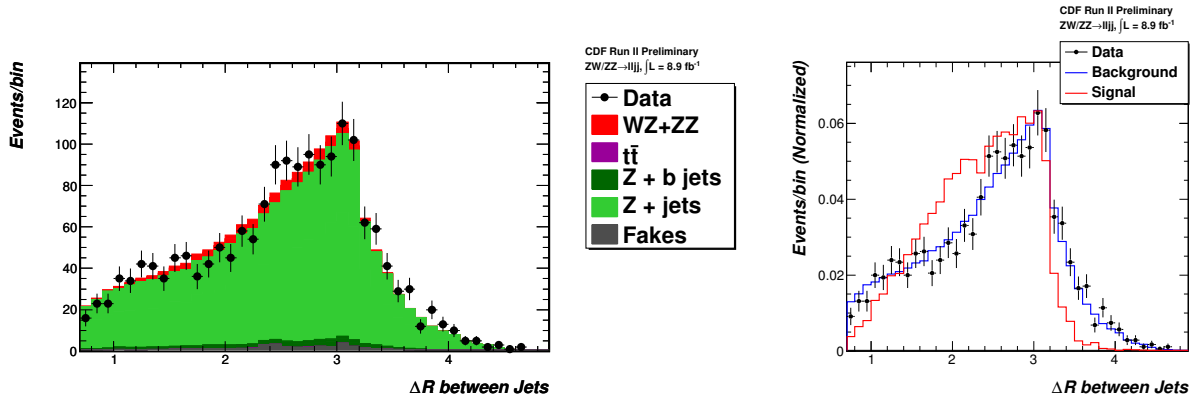

Figure B.10: The distribution of ΔR between jets in the Z + 2 jet signal region, with MC scaled to the data's luminosity (left) and with the distributions normalized to unity (right).

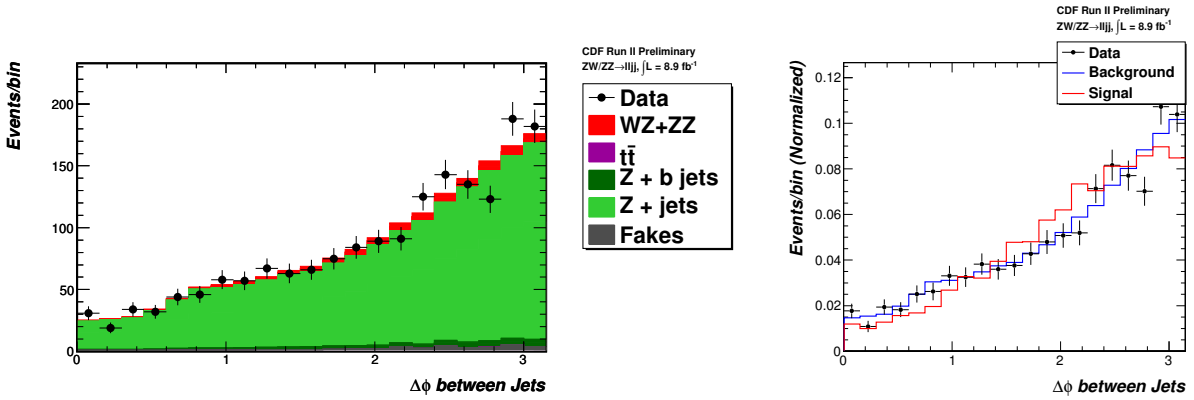

Figure B.11: The distribution of $\Delta\phi$ between jets in the Z + 2 jet signal region, with MC scaled to the data's luminosity (left) and with the distributions normalized to unity (right).

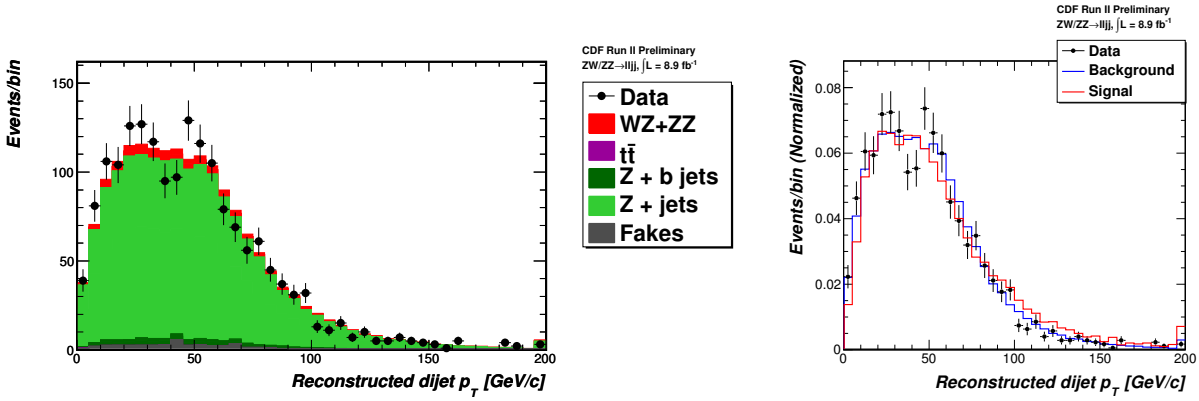

Figure B.12: The distribution of dijet p_T in the Z + 2 jet signal region, with MC scaled to the data's luminosity (left) and with the distributions normalized to unity (right).

B.3 Unagged Events

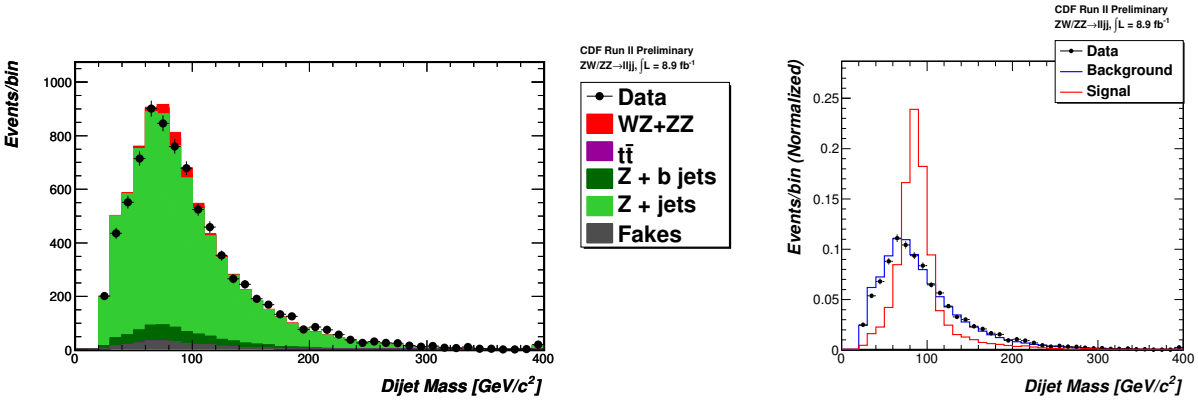

Figure B.13: The distribution of dijet mass in the Z + 2 jet signal region, with MC scaled to the data's luminosity (left) and with the distributions normalized to unity (right).

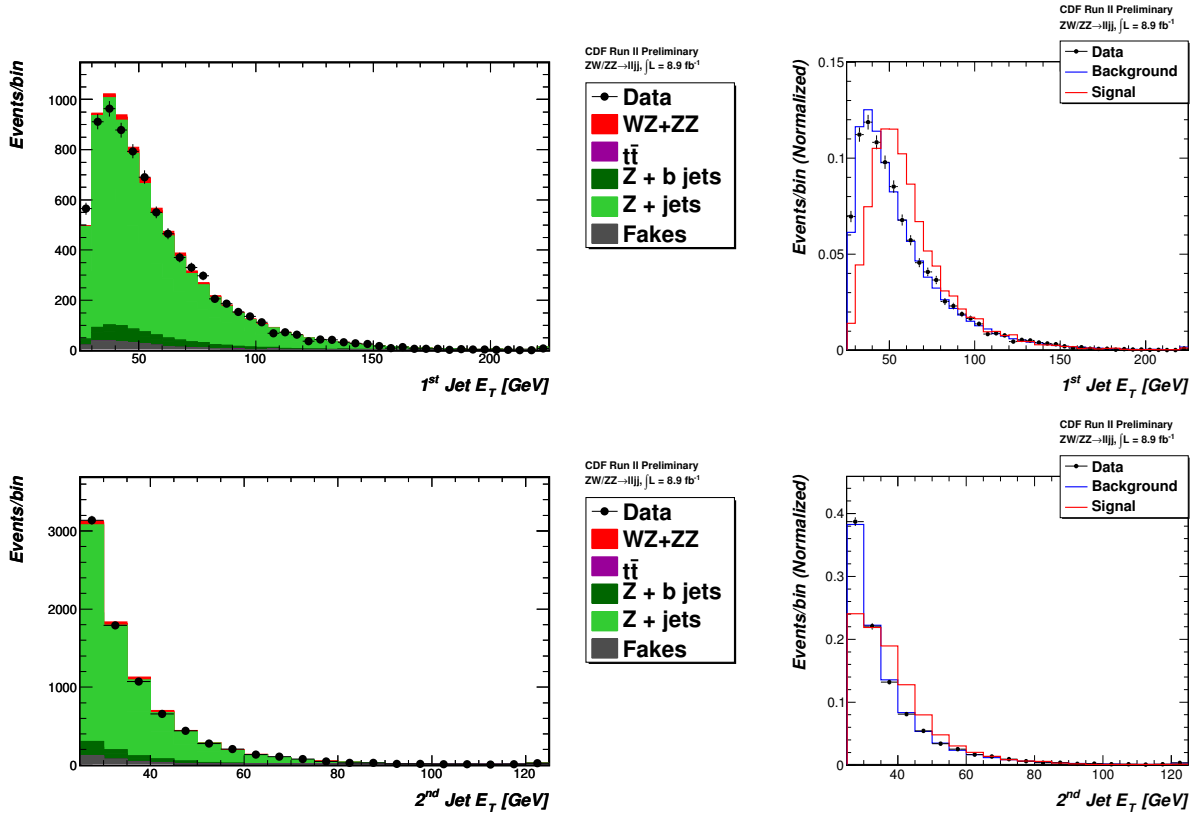

Figure B.14: The distribution of the leading jet E_T in the Z + 2 jet signal region, with MC scaled to the data's luminosity (top left) and with the distributions normalized to unity (top right), and the second leading jet E_T (bottom row).

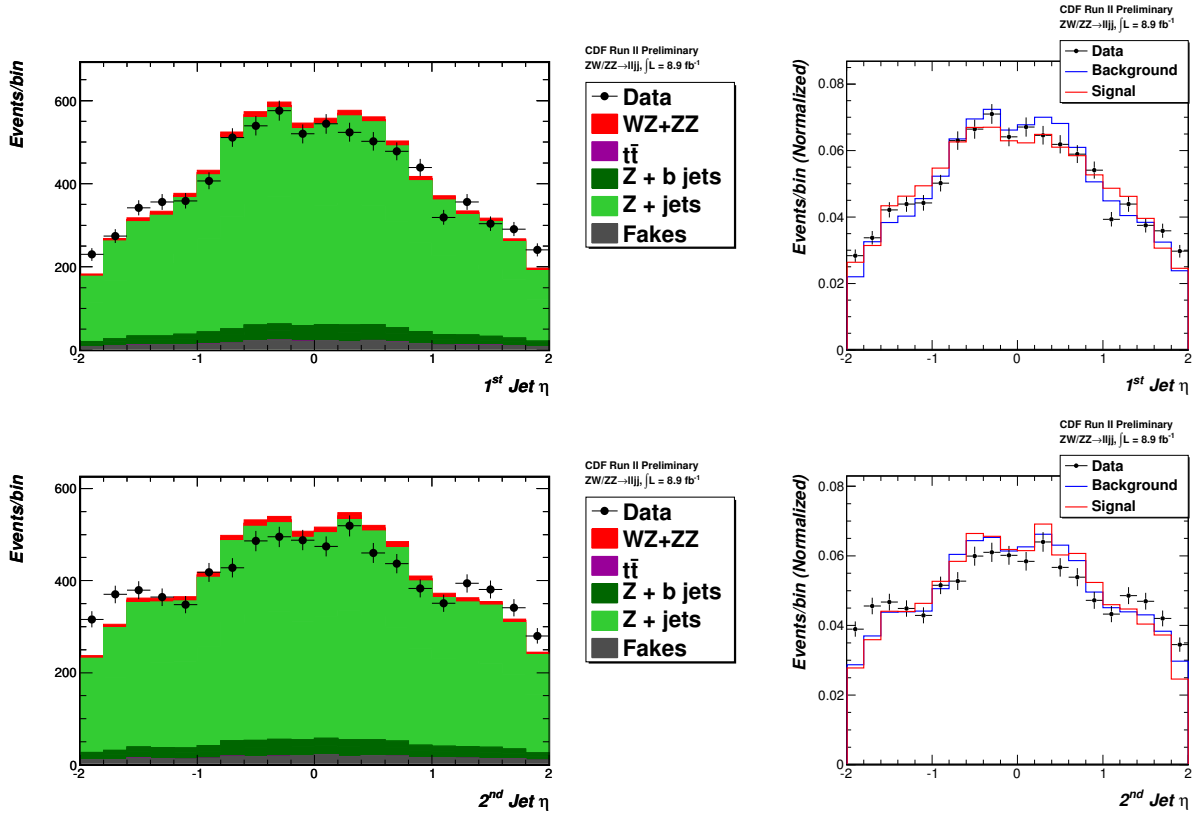

Figure B.15: The distribution of the leading jet η in the Z + 2 jet signal region, with MC scaled to the data's luminosity (top left) and with the distributions normalized to unity (top right), and the second leading jet η (bottom row).

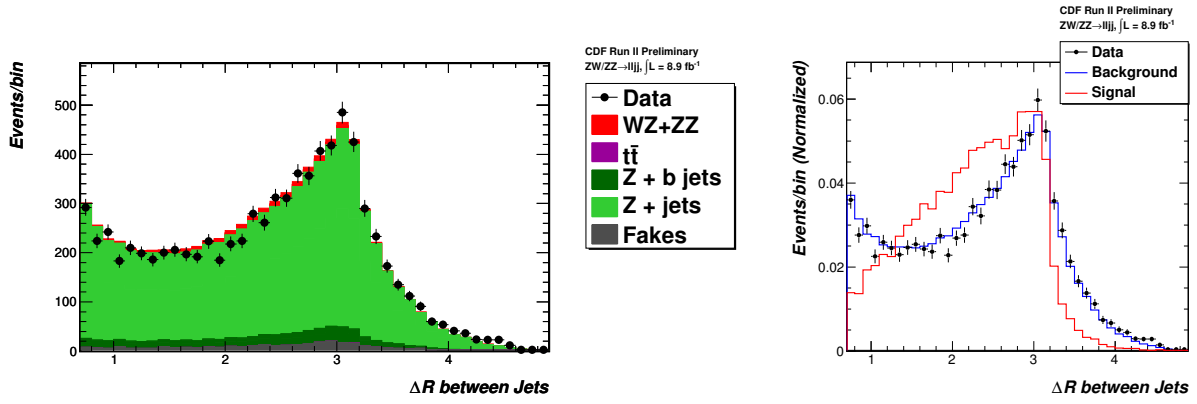

Figure B.16: The distribution of ΔR between jets in the Z + 2 jet signal region, with MC scaled to the data's luminosity (left) and with the distributions normalized to unity (right).

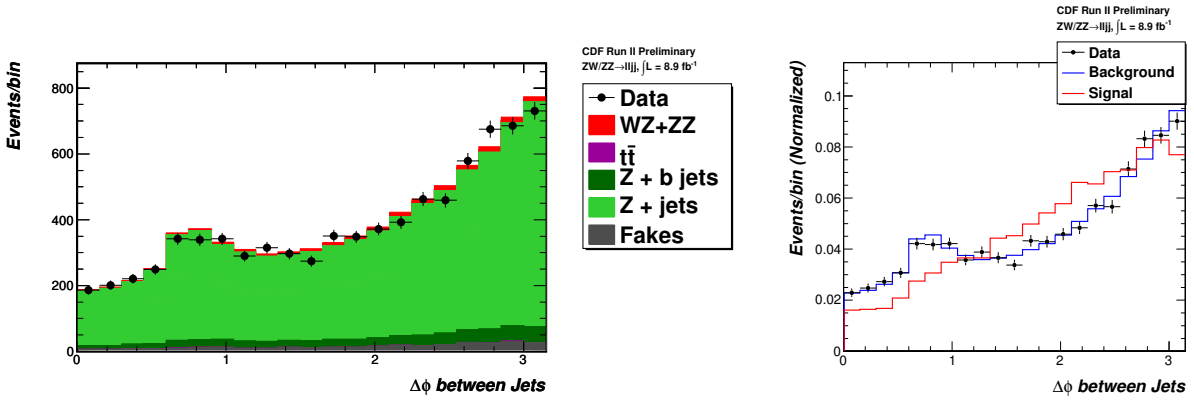

Figure B.17: The distribution of $\Delta\phi$ between jets in the Z + 2 jet signal region, with MC scaled to the data's luminosity (left) and with the distributions normalized to unity (right).

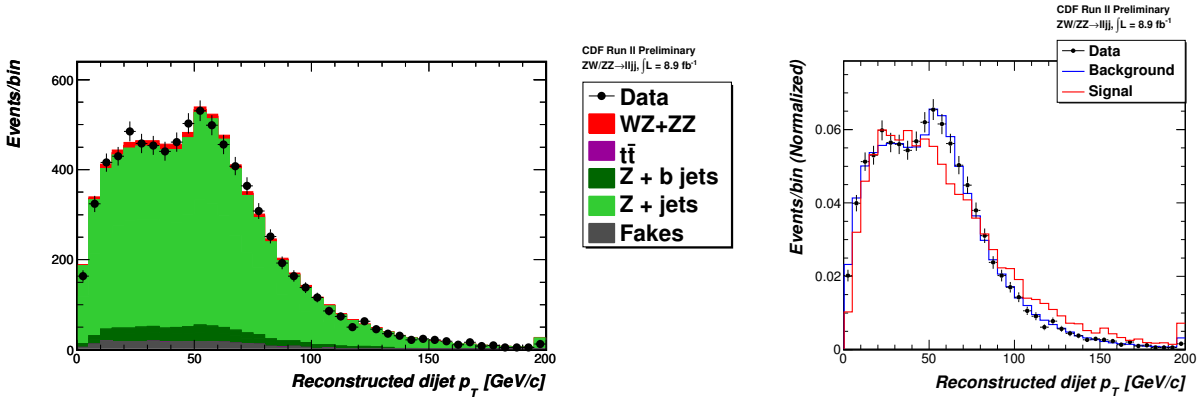

Figure B.18: The distribution of dijet p_T in the Z + 2 jet signal region, with MC scaled to the data's luminosity (left) and with the distributions normalized to unity (right).

APPENDIX C

BACKGROUND SYSTEMATIC SHAPE UNCERTAINTIES

The following plots show the histograms used as bounds on the systematic uncertainties in our Z + jets and Z + b jets templates. We show only the uncertainties that also have a shape uncertainty associated with them, but we show the normalization differences as well in these templates. These histograms are used as the inputs into `mclimit`, and are used to histograms for various values of the nuisance parameters through interpolation and extrapolation, described in Sec. 8.2.

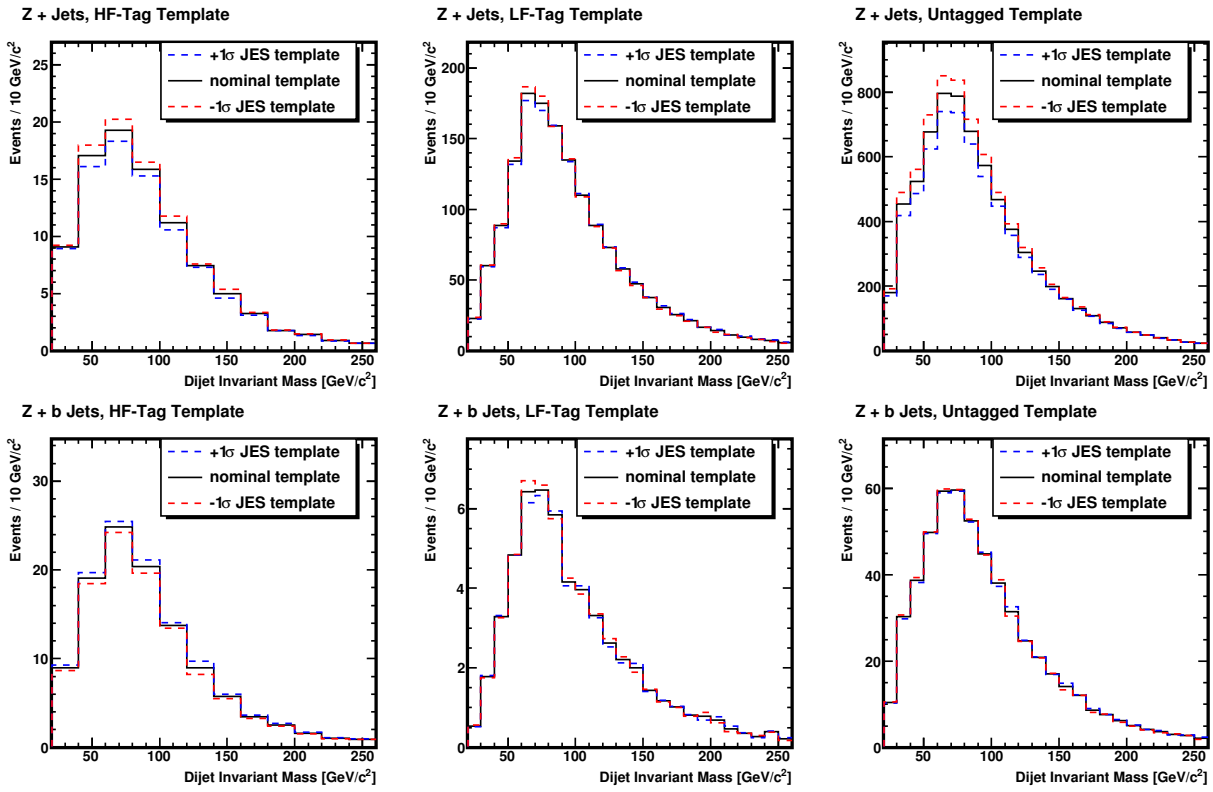

Figure C.1: The effect of the jet energy scale uncertainty on the Z + jets (top) and Z + b jets (bottom) backgrounds.

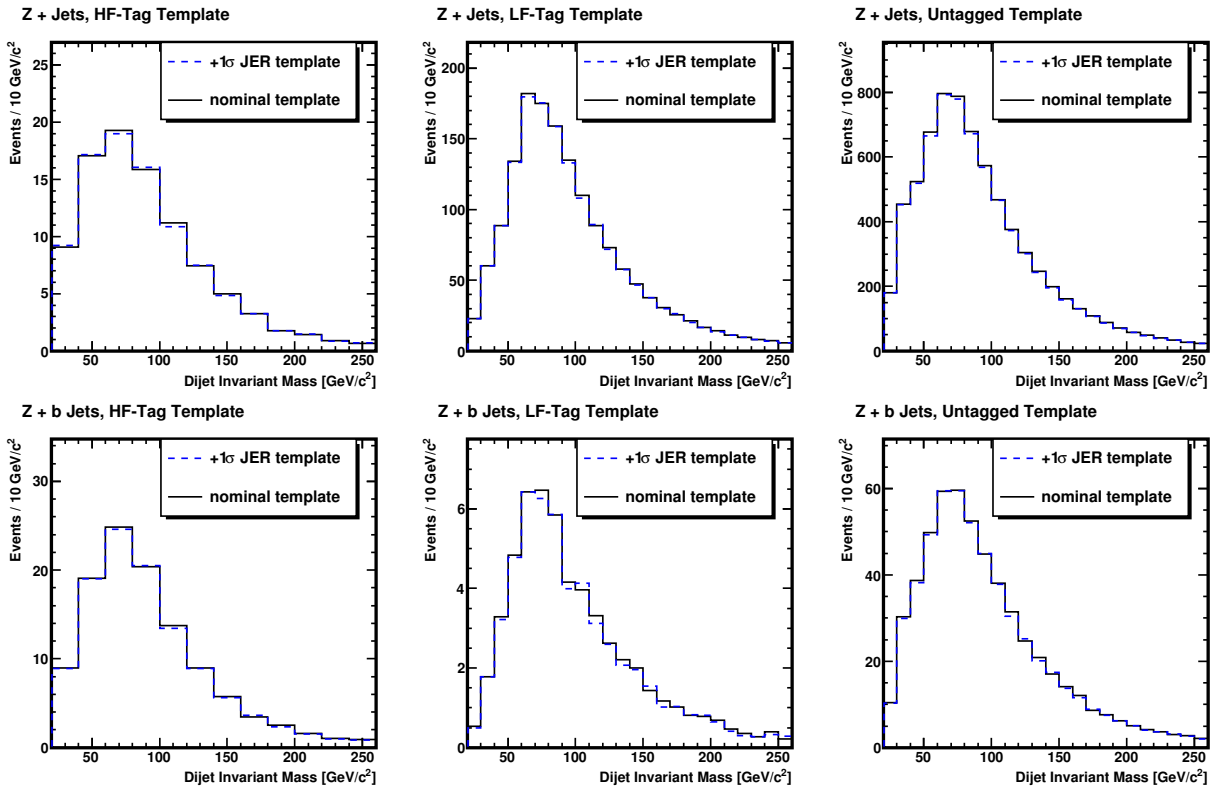

Figure C.2: The effect of the jet energy resolution uncertainty on the Z + jets (top) and Z + b jets (bottom) backgrounds.

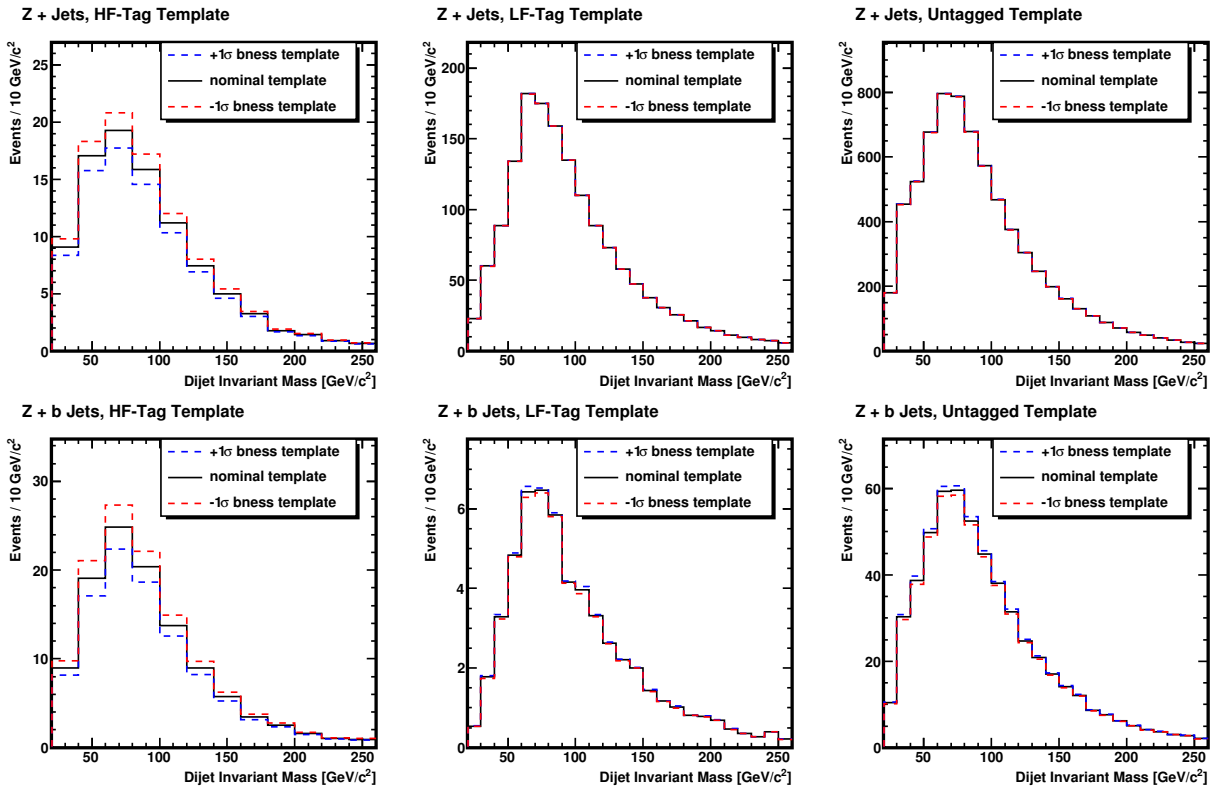

Figure C.3: The effect of the jet *b*ness tag uncertainty on the Z + jets (top) and $Z + b$ jets (bottom) backgrounds.

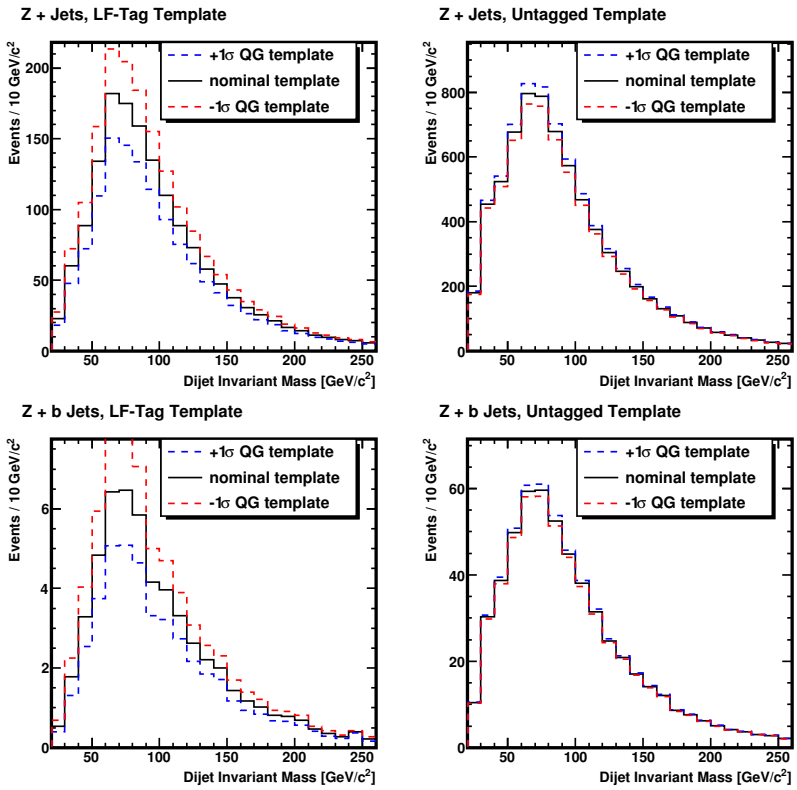

Figure C.4: The effect of the jet QG-value tag uncertainty on the Z + jets (top) and Z + b jets (bottom) backgrounds.

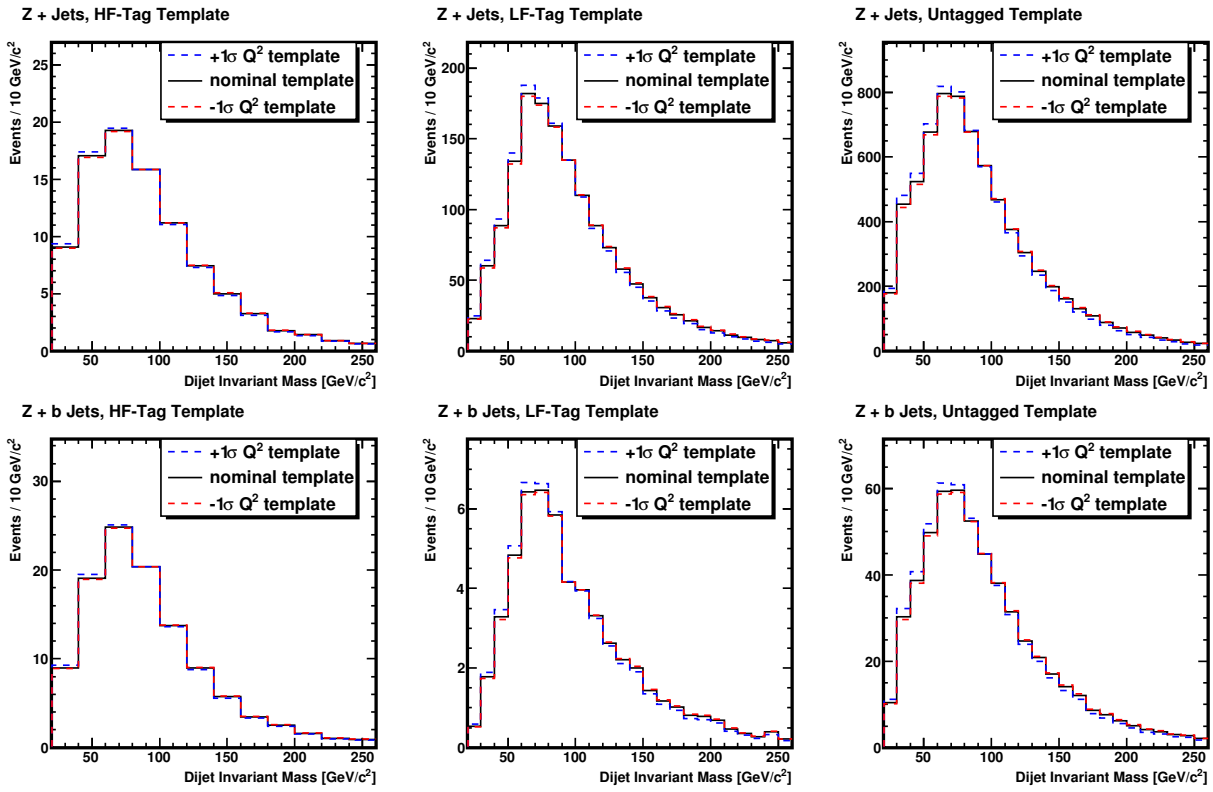

Figure C.5: The effect of the Q^2 parameter uncertainty on the Z + jets (top) and Z + b jets (bottom) backgrounds.

APPENDIX D

SIGNAL SYSTEMATIC SHAPE UNCERTAINTIES

The following plots show the fits to the distributions of our various shape uncertainties on the signal shape. We use the results of these fits to modify the signal shape inside `mclimit`. See Section 8.2 for details.

D.1 Jet Energy Scale

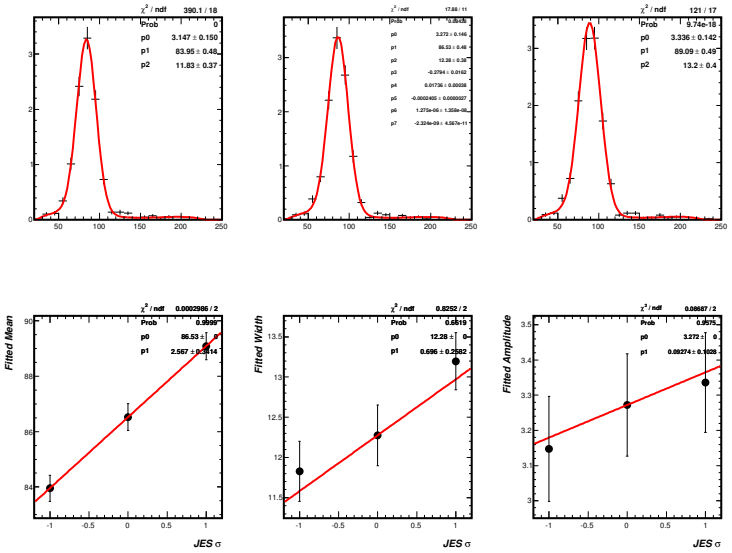

Figure D.1: Fits to the signal templates in channel 1 (HF tag), varying the jet energy scale. The top panel shows the Gaussian+fourth-degree polynomial fit to the dijet mass distributions. The fourth-degree polynomial is fit in the central template, and then remains fixed in the other templates. The bottom panel shows the Gaussian parameters of the fit as a function of the JES sigma. A line fits these points very well, and we use that fit to form templates for any value of JES required in mclimit.

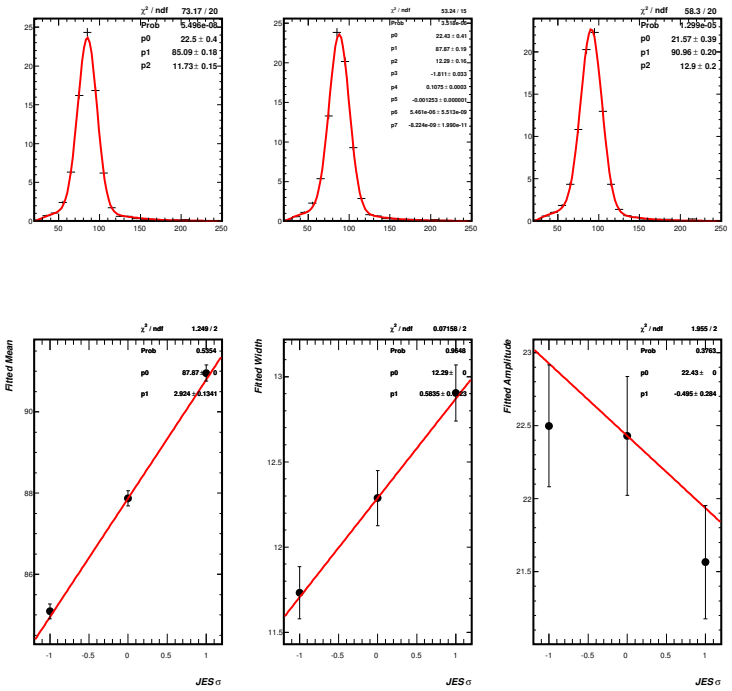

Figure D.2: Fits to the signal templates in channel 2 (LF tag), varying the jet energy scale. See the caption in Figure D.1 for a detailed description.

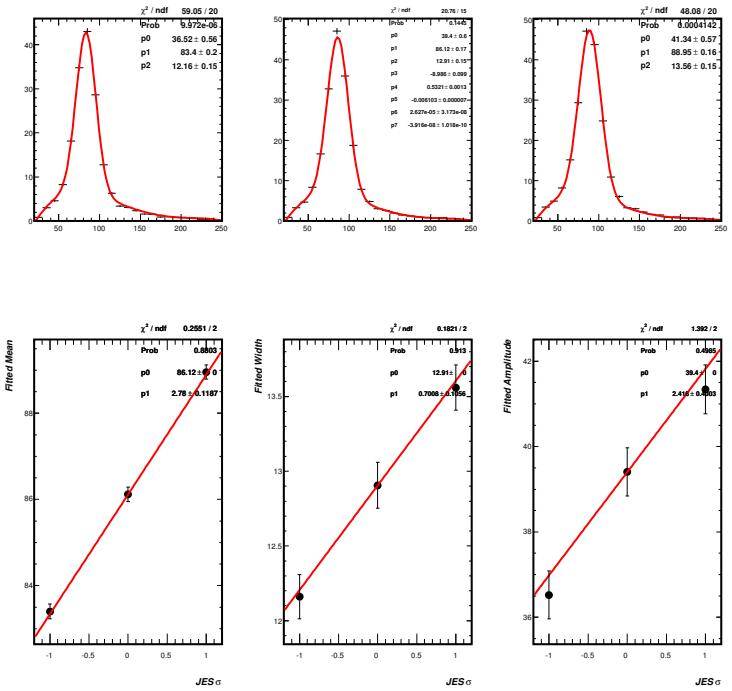

Figure D.3: Fits to the signal templates in channel 3 (no tag), varying the jet energy scale. See the caption in Figure D.1 for a detailed description.

D.2 Jet Energy Resolution

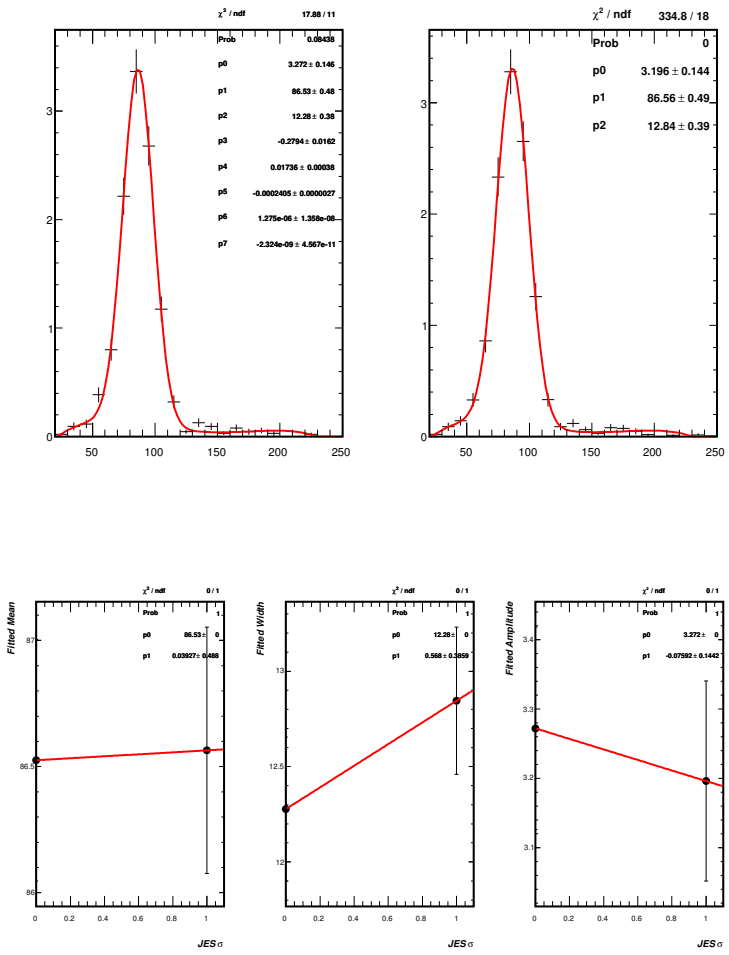

Figure D.4: Fits to the signal templates in channel 1 (HF tag), varying the jet resolution.

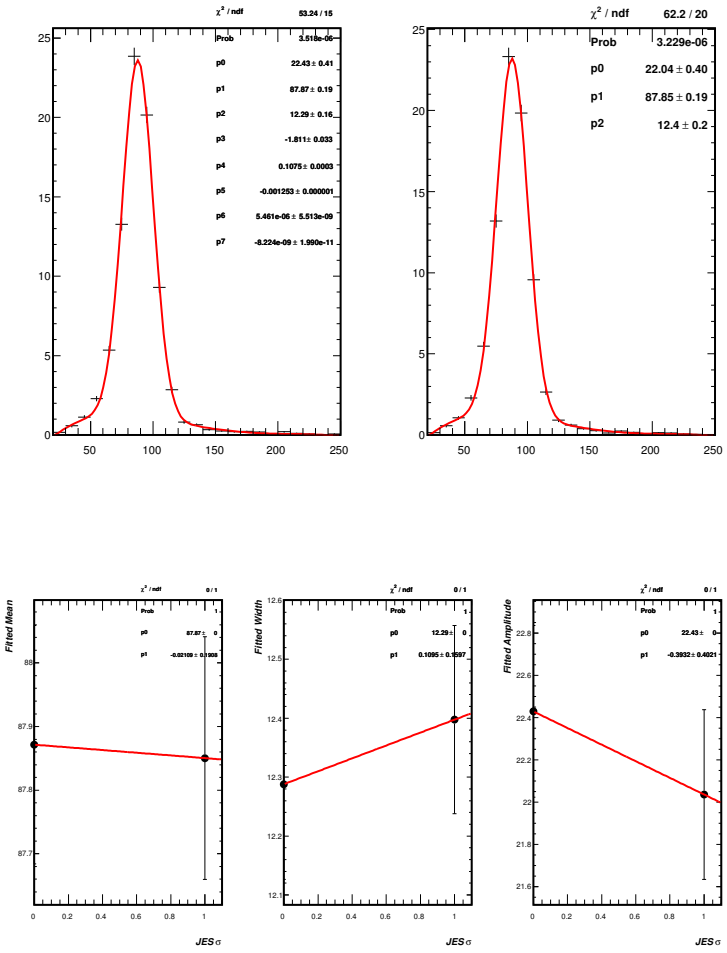

Figure D.5: Fits to the signal templates in channel 2 (LF tag), varying the jet resolution. See the caption in Figure D.1 for a detailed description.

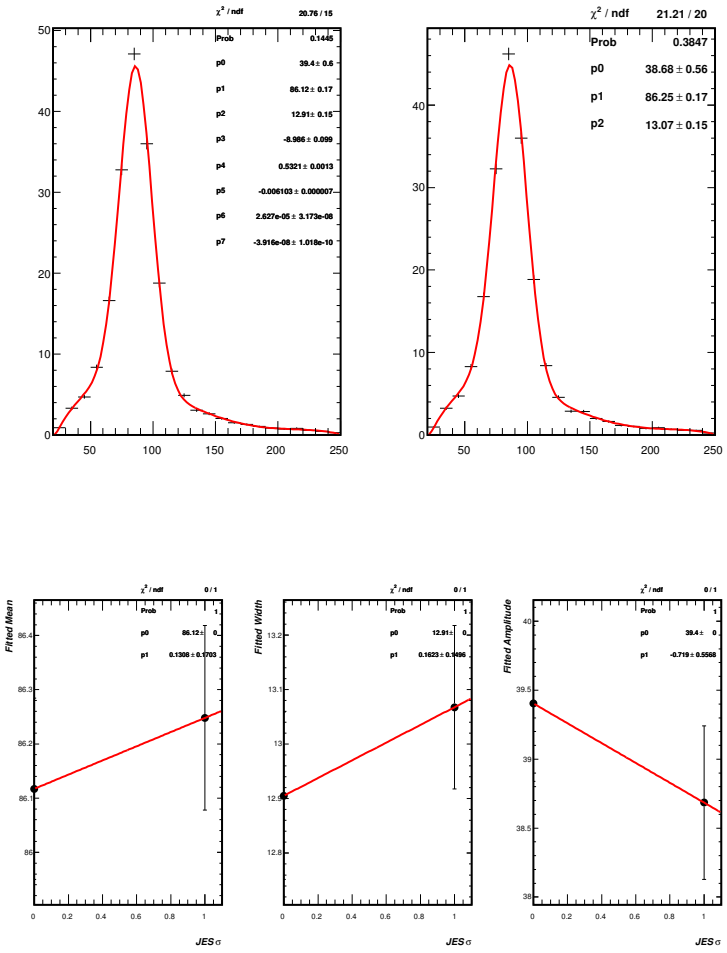

Figure D.6: Fits to the signal templates in channel 3 (no tag), varying the jet resolution. See the caption in Figure D.1 for a detailed description.

APPENDIX E

CHECKS ON FITTING PROCEDURE VIA PSEUDO-DATA FITS

Here we show the fit values from pseudo-data drawn according to our central value templates. The actual values returned from the fit for the various parameters are shown on the left, with the difference between the thrown and fit value on the right. We expect to see a mean of 0 and width of 1 for the latter pull distributions. The units are in the σ of the parameter, except in the case of the Z + jets and signal normalizations, which are in fractions of the total normalization.

We see that the means are broadly consistent with 0σ, indicating that our fits are unbiased. There are a few parameters that differ from 0, like the jet energy scale an Q^2, but these deviations are on the order of a few percent of a σ, and thus are not cause for major concern. They are highly correlated with each other and with changes in the Z + jets normalizations, which is adjusted in these pseudo-experiments to match the normalization determined from our fit to the data. The *b*ness parameter also shows a strange asymmetric shape, but this is consistent with a close correlation to the $Z + b\bar{b}$ normalization, which is not allowed to go below zero and thus puts a bound on movement of the *b*ness parameter.

For some parameters, the widths of the fit parameter distributions are smaller than 1, indicating that our fit has greater sensitivity to changes in the fit parameters than the input systematic uncertainty. Sensitivity studies are done with the full input systematic uncertainties, so we are being conservative when it comes to our treatment of these parameters.

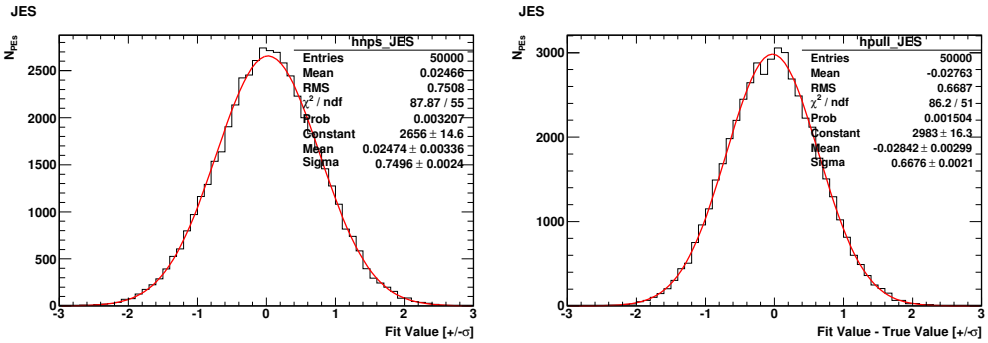

Figure E.1: Fit (*left*) and pull (*right*) for the jet energy scale parameter.

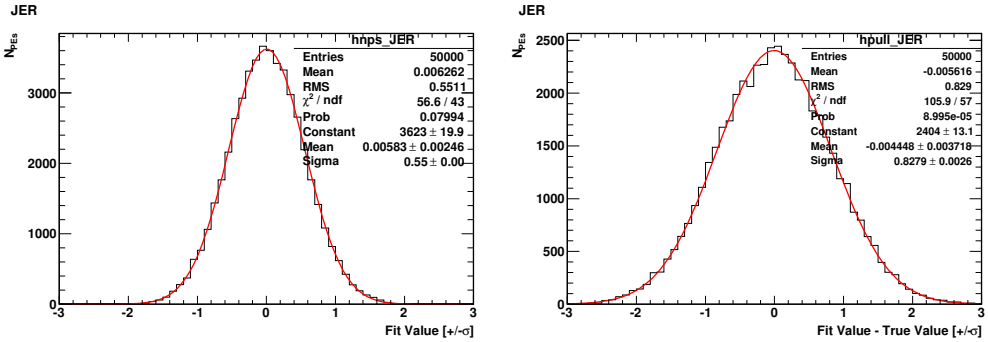

Figure E.2: Fit (*left*) and pull (*right*) for the jet energy resolution parameter.

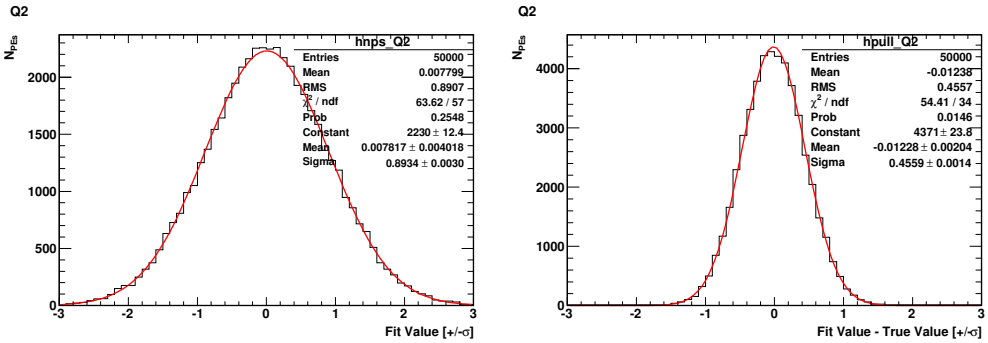

Figure E.3: Fit (*left*) and pull (*right*) for the Q^2 parameter.

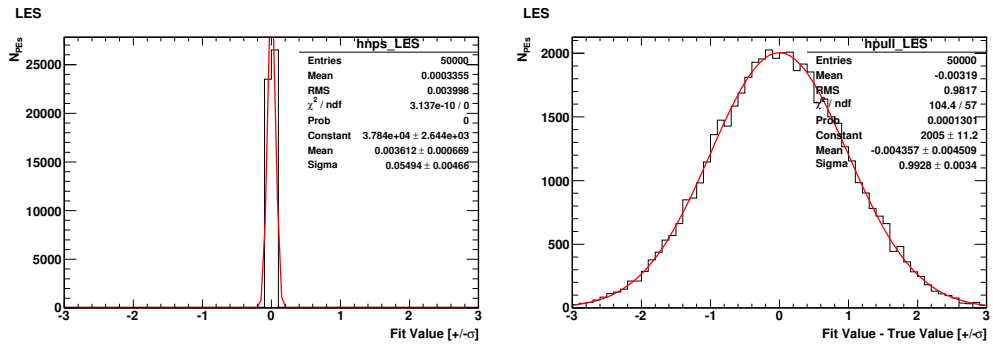

Figure E.4: Fit (*left*) and pull (*right*) for the lepton energy scale parameter.

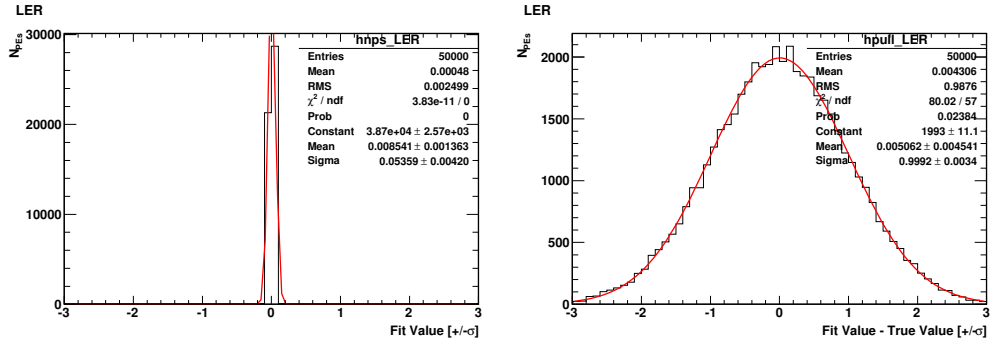

Figure E.5: Fit (*left*) and pull (*right*) for the lepton energy resolution parameter.

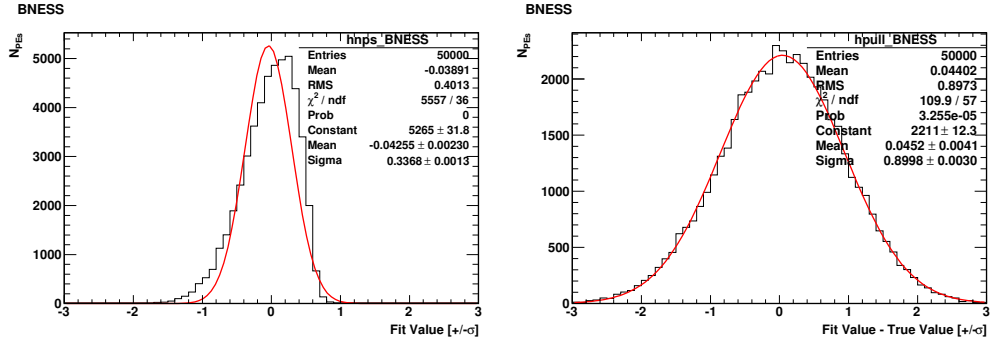

Figure E.6: Fit (*left*) and pull (*right*) for the *b*ness parameter.

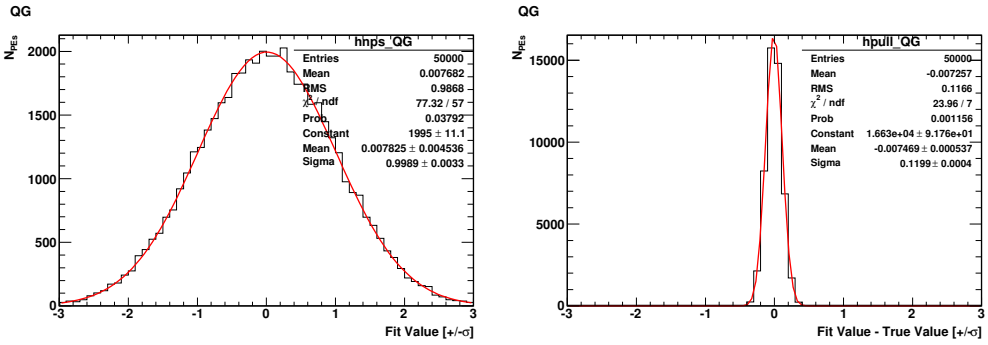

Figure E.7: Fit (*left*) and pull (*right*) for the QG value parameter.

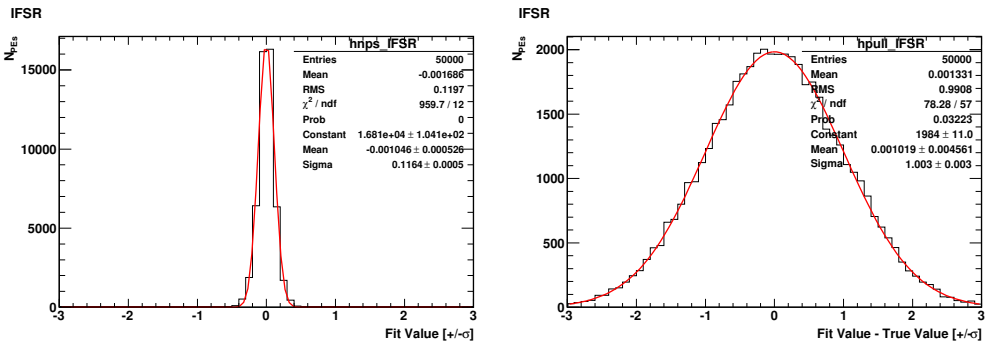

Figure E.8: Fit (*left*) and pull (*right*) for the initial/final state radiation parameter.

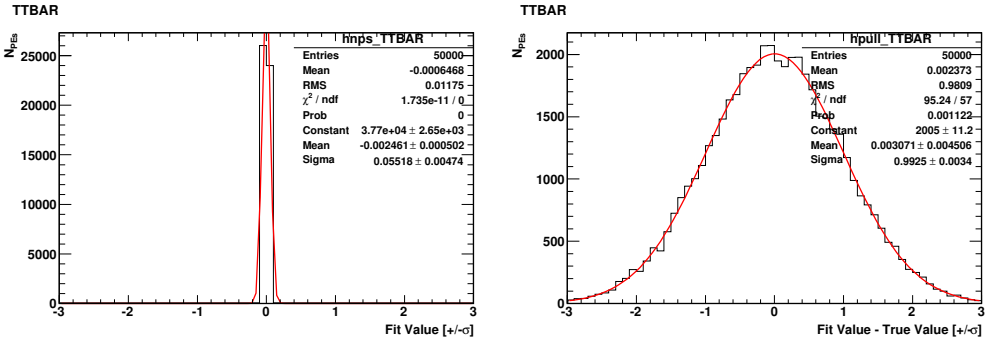

Figure E.9: Fit (*left*) and pull (*right*) for the $t\bar{t}$ jets cross section.

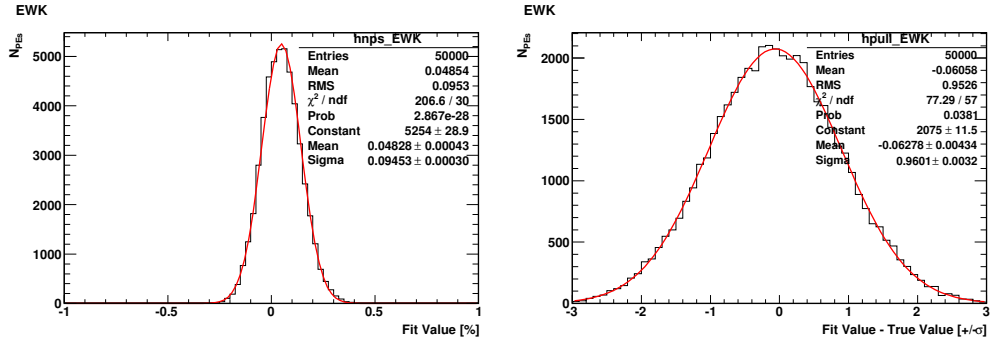

Figure E.10: Fit (*left*) and pull (*right*) for the Z + jets normalization, in units of %.

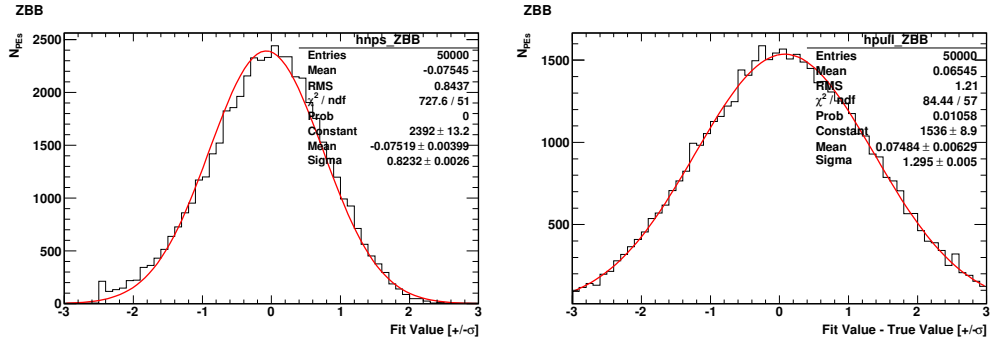

Figure E.11: Fit (*left*) and pull (*right*) for the Z + b jets normalization relative to the Z + jets normalization.

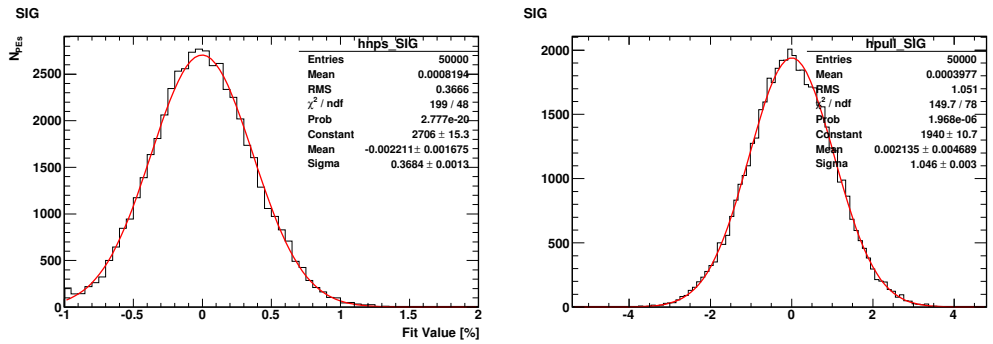

Figure E.12: Fit (*left*) and pull (*right*) for the ZW/ZZ normalization, in units of %.

APPENDIX F

ADDITIONAL FITS FOR SIGNAL

We include a number of fits to subsets of our selection, for checks on consistency. We show both the final fit, as well as the final fit parameters. In the fit with signal constrained, Sec. F.7, we constrain the signal cross section in the templates to $\pm$6%.

A discussion of some of the conclusions from these fits is provided in Sec. 8.3.1.

F.1 Electrons Only Fit

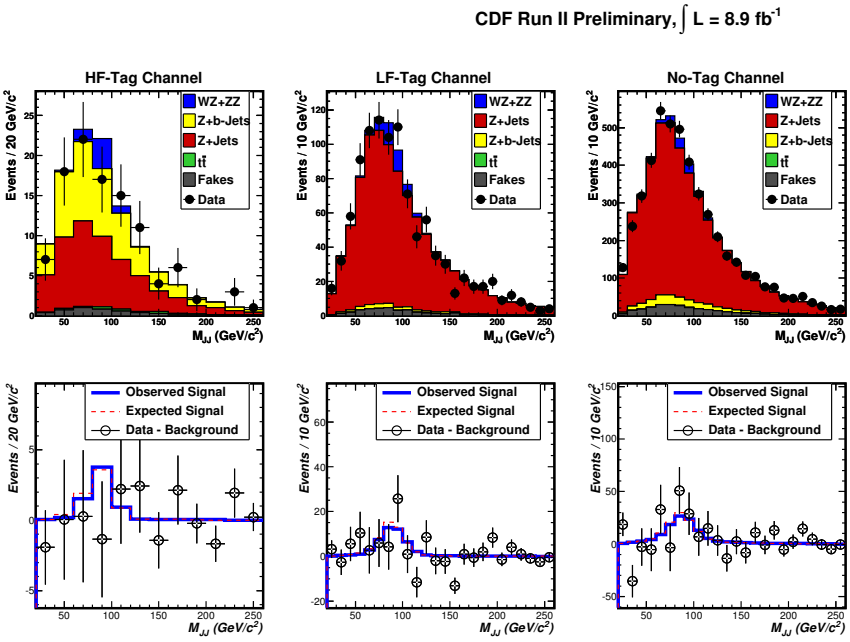

Figure F.1: Fit to events with electron events.

Template	Fit Normalization	$\pm$	Error
Z+jets	+2.9%	$\pm$	5.6%
$Z + b\bar{b}$	$-$28%	$\pm$	24%
$t\bar{t}$	+0.0%	$\pm$	6.5%
Fakes	$-$6%	$\pm$	49%
ZW/ZZ	$-$7%	$\pm$	46%

Systematic	Fit Value (σ)	$\pm$	Error (σ)
Jet Resolution	+0.25	$\pm$	0.85
Jet Energy Scale	+0.73	$\pm$	0.77
Q^2	$-$1.30	$\pm$	0.55
*b*ness Tag/Mistag	+0.37	$\pm$	0.97
QG Tag/Mistag	+0.13	$\pm$	0.14
Lepton Energy Scale	0.00	$\pm$	1.00
Lepton Energy Resolution	0.00	$\pm$	1.00
ISR/FSR	$-$0.02	$\pm$	1.00

Table F.1: Fit parameter results with only electron events.

F.2 Muons Only Fit

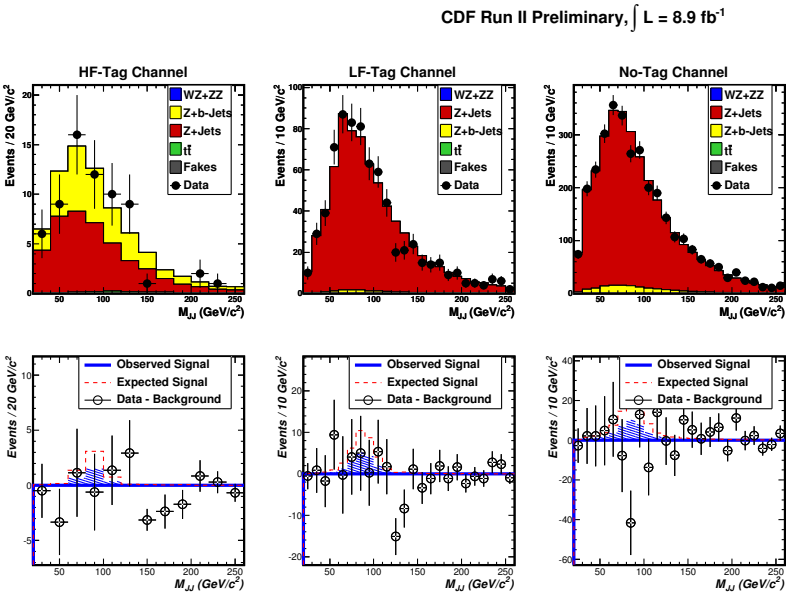

Figure F.2: Fit to events with muon events.

Template	Fit Normalization	±	Error
Z+jets	+10.0%	±	5.9%
$Z + b\bar{b}$	−43%	±	24%
$t\bar{t}$	−0.1%	±	6.5%
Fakes	−3%	±	50%
ZW/ZZ	−100%	+	42%
		−	0% (*at limit*)

Systematic	Fit Value (σ)	±	Error (σ)
Jet Resolution	+0.17	±	0.89
Jet Energy Scale	+0.61	±	0.83
Q^2	−0.73	±	0.55
*b*ness Tag/Mistag	+0.61	±	1.03
QG Tag/Mistag	−0.17	±	0.16
Lepton Energy Scale	−0.00	±	1.00
Lepton Energy Resolution	−0.00	±	1.00
ISR/FSR	+0.00	±	1.00

Table F.2: Fit parameter results with only muon events.

F.3 Fit Excluding Muon Untagged Channel

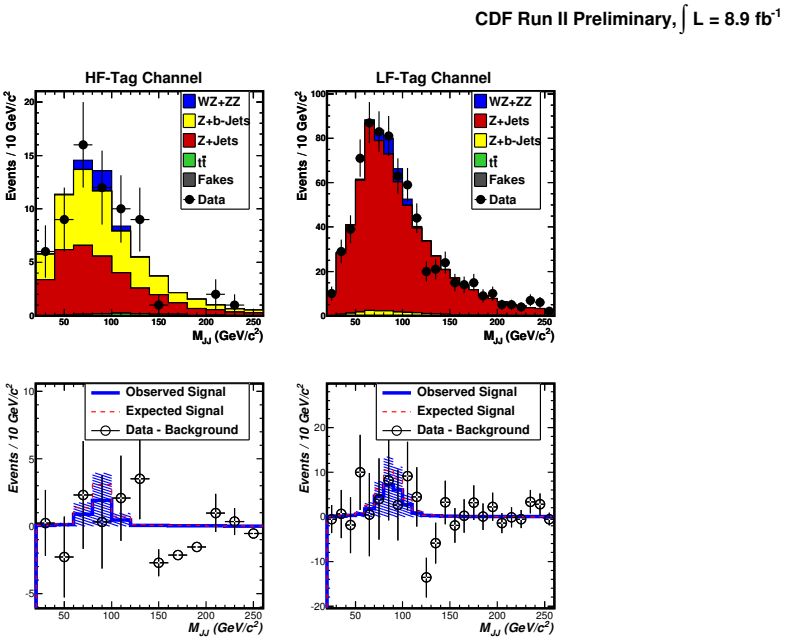

Figure F.3: Fit when excluding the untagged channel in muon events.

Template	Fit Normalization	$\pm$	Error
Z+ jets	+2.6%	$\pm$	5.3%
$Z + b\bar{b}$	$-$38%	$\pm$	20%
$t\bar{t}$	$-$0.0%	$\pm$	6.5%
Fakes	+5%	$\pm$	46%
ZW/ZZ	$-$13%	$\pm$	38%

Systematic	Fit Value (σ)	$\pm$	Error (σ)
Jet Resolution	+0.27	$\pm$	0.85
Jet Energy Scale	+0.44	$\pm$	0.75
Q^2	$-$1.29	$\pm$	0.54
*b*ness Tag/Mistag	+0.51	$\pm$	0.98
QG Tag/Mistag	$-$0.03	$\pm$	0.14
Lepton Energy Scale	+0.01	$\pm$	1.00
Lepton Energy Resolution	+0.00	$\pm$	1.00
ISR/FSR	+0.00	$\pm$	1.00

Table F.3: Fit parameter results excluding the untagged channel in muon events.

F.4 Fit in HF-Tag Channel Only

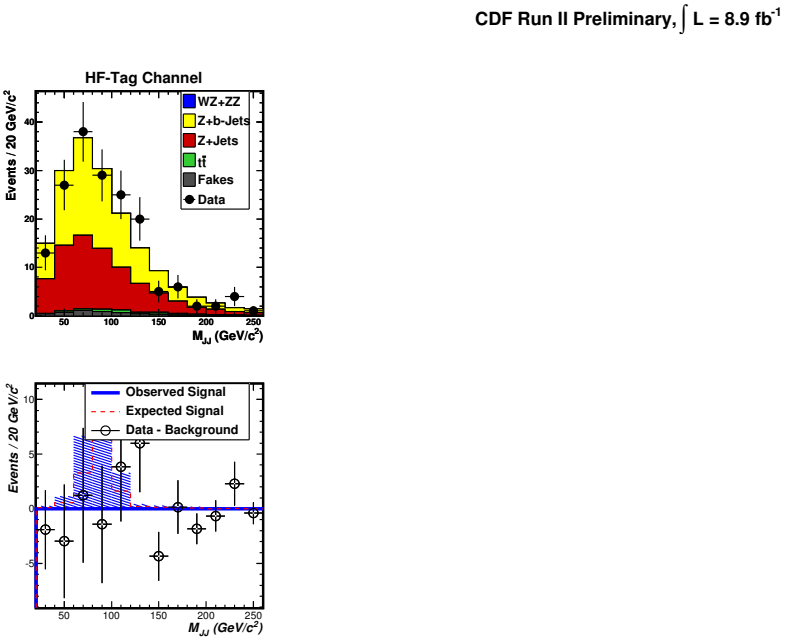

Figure F.4: Fit for only HF-Tag region.

Template	Fit Normalization	±	Error
Z+ jets	−20%	±	24%
$Z + b\bar{b}$	+2%	±	47%
$t\bar{t}$	+0.0%	±	6.5%
Fakes	−0%	±	50%
ZW/ZZ	−100%	+	200%
		-	0% (*at limit*)

Systematic	Fit Value (σ)	±	Error (σ)
Jet Resolution	−0.06	±	0.98
Jet Energy Scale	+0.12	±	0.98
Q^2	−0.10	±	0.98
*b*ness Tag/Mistag	0.07	±	1.01
Lepton Energy Scale	0.00	±	1.00
Lepton Energy Resolution	0.00	±	1.00
ISR/FSR	0.00	±	1.00

Table F.4: Fit parameter results from fit to HF-Tag only region.

F.5 Fit in LF-Tag Channel Only

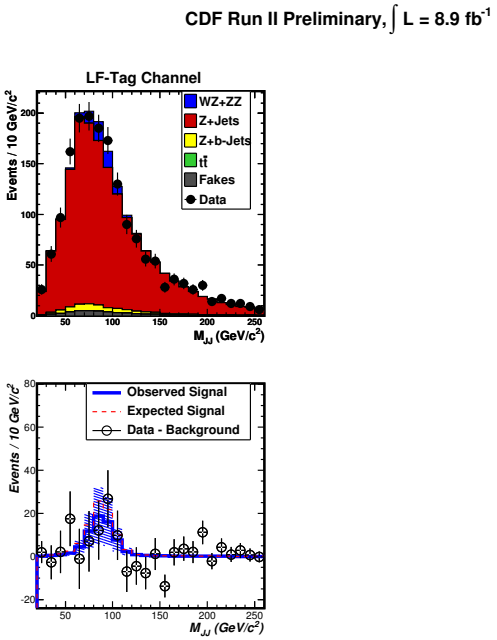

Figure F.5: Fit for only LF-Tag region.

Template	Fit Normalization	$\pm$	Error
Z+ jets	+2.0%	$\pm$	32%
$Z + b\bar{b}$	+0.64%	$\pm$	40%
$t\bar{t}$	+0.0%	$\pm$	6.5%
Fakes	+2.1%	$\pm$	50%
ZW/ZZ	-26%	$\pm$	48%

Systematic	Fit Value (σ)	$\pm$	Error (σ)
Jet Resolution	-0.09	$\pm$	0.90
Jet Energy Scale	-0.04	$\pm$	0.79
Q^2	-0.47	$\pm$	0.70
*b*ness Tag/Mistag	-0.01	$\pm$	0.99
QG Tag/Mistag	-0.02	$\pm$	0.95
Lepton Energy Scale	0.00	$\pm$	1.00
Lepton Energy Resolution	0.00	$\pm$	1.00
ISR/FSR	-0.01	$\pm$	0.99

Table F.5: Fit parameter results from fit to LF-Tag only region.

F.6 Fit in Untagged Channel Only

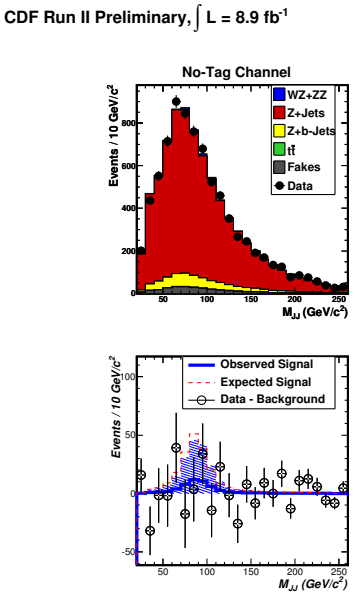

Figure F.6: Fit for only untagged region.

Template	Fit Normalization	±	Error
Z+ jets	+5.4%	±	10%
$Z + b\bar{b}$	+1.2%	±	43%
$t\bar{t}$	+0.0%	±	6.5%
Fakes	+2.9%	±	49%
ZW/ZZ	−76%	+	101%
		−	34%

Systematic	Fit Value (σ)	±	Error (σ)
Jet Resolution	+0.26	±	0.94
Jet Energy Scale	+1.02	±	0.81
Q^2	−1.08	±	0.51
*b*ness Tag/Mistag	+0.00	±	1.00
QG Tag/Mistag	+0.06	±	0.99
Lepton Energy Scale	0.00	±	1.00
Lepton Energy Resolution	0.00	±	1.00
ISR/FSR	−0.00	±	1.00

Table F.6: Fit parameter results from fit to untagged only region.

F.7 Fit with Signal Constrained

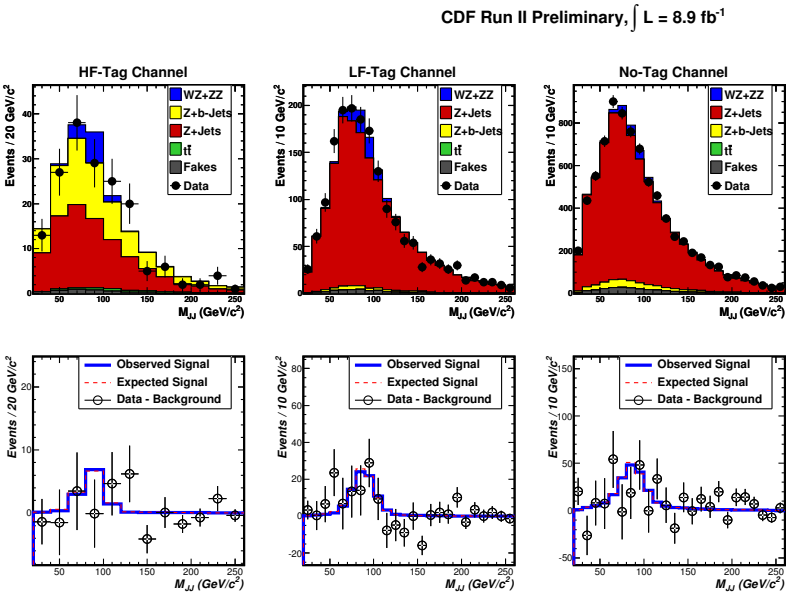

Figure F.7: Fit with signal constrained.

Template	Fit Normalization	$\pm$	Error
Z+ jets	+3.1%	$\pm$	4.8%
$Z + b\bar{b}$	-39%	$\pm$	19%
$t\bar{t}$	-0.0%	$\pm$	6.5%
Fakes	-9%	$\pm$	48%
ZW/ZZ	-1.5%	$\pm$	5.9%
Systematic	Fit Value (σ)	$\pm$	Error (σ)
Jet Resolution	+0.14	$\pm$	0.84
Jet Energy Scale	+0.34	$\pm$	0.74
Q^2	-1.35	$\pm$	0.49
*b*ness Tag/Mistag	+0.53	$\pm$	0.94
QG Tag/Mistag	-0.01	$\pm$	0.12
Lepton Energy Scale	-0.02	$\pm$	1.00
Lepton Energy Resolution	+0.00	$\pm$	1.00
ISR/FSR	-0.07	$\pm$	1.00
χ^2/d.o.f, No Signal	62.0 / 56		
χ^2/d.o.f, w/ Signal	61.8 / 56		

Table F.7: Fit parameter results from fit with signal constrained.

REFERENCES

[1] Harris, D. and K. Riesselmann. "Deconstruction: Standard Model Discoveries." Symmetry Magazine **6** (2009) 30–31.

[2] Englert, F. and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons." Phys.Rev.Lett. **13** (1964) 321–323.

[3] Higgs, P. W. "Broken Symmetries and the Masses of Gauge Bosons." Phys.Rev.Lett. **13** (1964) 508–509.

[4] Guralnik, G., C. Hagen, and T. Kibble. "Global Conservation Laws and Massless Particles." Phys.Rev.Lett. **13** (1964) 585–587.

[5] Glashow, S. "Partial Symmetries of Weak Interactions." Nucl.Phys. **22** (1961) 579–588.

[6] Weinberg, S. "A Model of Leptons." Phys.Rev.Lett. **19** (1967) 1264–1266.

[7] Salam, A. *in Elementary Particle Theory, p. 367*. Almqvist and Wiksell, Stockholm (1968).

[8] Hasert, F. et al. "Observation of Neutrino Like Interactions Without Muon Or Electron in the Gargamelle Neutrino Experiment." Phys.Lett. **B46** (1973) 138–140.

[9] Arnison, G. et al. "Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at s**(1/2) = 540-GeV." Phys.Lett. **B122** (1983) 103–116.

[10] Arnison, G. et al. "Experimental Observation of Lepton Pairs of Invariant Mass Around 95-GeV/c**2 at the CERN SPS Collider." Phys.Lett. **B126** (1983) 398–410.

[11] Banner, M. et al. "Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN anti-p p Collider." Phys.Lett. **B122** (1983) 476–485.

[12] Bagnaia, P. et al. "Evidence for $Z^0 \to e^+e^-$ at the CERN anti-p p Collider." Phys.Lett. **B129** (1983) 130–140.

[13] Baak, M., et al. "Updated Status of the Global Electroweak Fit and Constraints on New Physics." Eur.Phys.J. **C72** (2012) 2003.

[14] Barate, R. et al. "Search for the standard model Higgs boson at LEP." Phys.Lett. **B565** (2003) 61–75.

[15] Aad, G. et al. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC." arXiv:1207.7214 [hep-ex].

[16] Chatrchyan, S. et al. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC." arXiv:1207.7235 [hep-ex].

[17] Aaltonen, T. et al. "Evidence for a particle produced in association with weak bosons and decaying to a bottom-antibottom quark pair in Higgs boson searches at the Tevatron." arXiv:1207.6436 [hep-ex].

[18] Beringer, J. and others (Particle Data Group). Phys. Rev. D **86** (2012) 010001.

[19] Campbell, J. M. and R. K. Ellis. "Update on vector boson pair production at hadron colliders." Phys. Rev. D **60** (1999) 113006.

[20] Hagiwara, K. et al. "Probing the weak boson sector in $e^+e^- \to W^+W^-$." Nucl. Phys. **B 282** (1987) 253.

[21] Kober, M., B. Koch, and M. Bleicher. "First order calculation of the inclusive cross section $pp \to ZZ$ by graviton exchange in large extra dimensions." Phys. Rev. D **76** (2007) 125001.

[22] Eichten, E. J., K. Lane, and A. Martin. "Technicolor Explanation for the CDF Wjj Excess." Phys. Rev. Lett. **106** (2011) 251803.

[23] Buckley, M. R., et al. "Light $Z^{'}$ bosons at the Tevatron." Phys. Rev. D **83** (2011) 115013.

[24] Aaltonen, T. et al. "Invariant Mass Distribution of Jet Pairs Produced in Association with a W Boson in $p\overline{p}$ Collisions at $\sqrt{s}$ = 1.96 TeV." Phys. Rev. Lett. **106** (2011) 171801.

[25] Abazov, V. M., et al. "Bounds on an Anomalous Dijet Resonance in W + jets Production in $p\overline{p}$ Collisions at $\sqrt{s}$ = 1.96 TeV." Phys. Rev. Lett. **107** (2011) 011804.

[26] "Study of the dijet mass spectrum in W+2 jets events." CMS-PAS-EWK-11-017 .

[27] "Invariant mass distribution of jet pairs produced in association with a leptonically decaying W boson using 1.02 fb-1 of ATLAS data." ATLAS-CONF-2011-097 .

[28] Dittmaier, S., et al. "Handbook of LHC Higgs Cross Sections: 2. Differential Distributions." arXiv:1201.3084 [hep-ph].

[29] Neubauer, M. S. "Diboson production at colliders." Ann.Rev.Nucl.Part.Sci. **61** (2011) 223–250.

[30] Campbell, J., R. Ellis, and C. Williams. "Vector boson pair production at the LHC." Journal of High Energy Physics **2011** (2011) 1–36.

[31] Aaltonen, T. et al. “First Observation of Vector Boson Pairs in a Hadronic Final State at the Tevatron Collider.” Phys. Rev. Lett. **103** (2009) 091803.

[32] Aaltonen, T. et al. “Search for WZ+ZZ production with missing transverse energy+jets with b enhancement at $\sqrt{s}$=1.96 TeV.” Phys. Rev. D **85** (2012) 012002.

[33] Aaltonen, T. et al. “Measurement of the WW+WZ Production Cross Section Using the Lepton+Jets Final State at CDF II.” Phys. Rev. Lett. **104** (2010) 101801.

[34] Abazov, V. M. et al. “Measurements of WW and WZ Production in W + jets Final States in $p\bar{p}$ Collisions.” Phys. Rev. Lett. **108** (2012) 181803.

[35] Aaltonen, T. et al. “Combined search for the standard model Higgs boson decaying to a bb pair using the full CDF data set.” `arXiv:1207.1707 [hep-ex]`.

[36] Abazov, V. M. et al. “Combined search for the standard model Higgs boson decaying to $b\bar{b}$ using the D0 Run II data set.” `arXiv:1207.6631 [hep-ex]`.

[37] Crawford, D. “Tevatron Rookie Book.” FNAL Accelerator Division. Online: http://www-bdnew.fnal.gov/operations/rookie_books/Tevatron_v2.3.pdf.

[38] Adapted from http://www.fnal.gov/pub/science/accelerator/. Accessed August 2012.

[39] From http://www.fnal.gov/pub/now/tevlum.html. Accessed August 2012.

[40] Abulencia, A. et al. “Measurements of inclusive W and Z cross sections in $p\bar{p}$ collisions at $\sqrt{s}$ = 1.96 TeV.” J. Phys. G **34** (2007) 2457.

[41] Sill, A. “CDF Run II silicon tracking projects.” Nucl. Instrum. Methods A **447**, 12 (2000) 1 – 8.

[42] Affolder, T. et al. “CDF Central Outer Tracker.” Nucl. Instrum. Methods A **526**, 3 (2004) 249 – 299.

[43] Ashmanskas, B. et al. “The CDF silicon vertex trigger.” Nucl.Instrum.Meth. **A518** (2004) 532–536.

[44] Adelman, J. A. et al. “The Silicon Vertex Trigger upgrade at CDF.” Nucl.Instrum.Meth. **A572** (2007) 361–364.

[45] Anikeev, K. et al. “CDF level 2 trigger upgrade.” IEEE Trans.Nucl.Sci. **53** (2006) 653–658.

[46] Abe, F. et al. “The CDF detector: an overview.” Nucl. Instrum. Methods A **271**, 3 (1988) 387 – 403.

[47] Abe, F. et al. "Topology of three-jet events in $p\bar{p}$ collisions at $\sqrt{s}$ = 1.8 TeV." Phys. Rev. D **45** (1992) 1448.

[48] Bhatti, A. et al. "Determination of the jet energy scale at the Collider Detector at Fermilab." Nucl. Instrum. Methods A **566**, 2 (2006) 375 – 412.

[49] Mangano, M. L. et al. "ALPGEN, a generator for hard multiparton processes in hadronic collisions." J. High Energy Phys. **07**, 07 (2003) 001.

[50] Sjöstrand, T. et al. "PYTHIA 6 physics and manual." J. High Energy Phys. **05** (2006) 026.

[51] Aaltonen, T. et al. "Measurement of cross sections for b jet production in events with a Z boson in $p\bar{p}$ collisions at $\sqrt{s}$ = 1.96 TeV." Phys. Rev. D **79** (2009) 052008.

[52] Pumplin, J. et al. "New Generation of Parton Distributions with Uncertainties from Global QCD Analysis." J. High Energy Phys. **0207** (2002) 012.

[53] Brun, R. et al. "GEANT3 manual." (1978). CERN Report CERN-DD-78-2-REV (unpublished).

[54] Grindhammer, G., M. Rudowicz, and S. Peters. "The fast simulation of electromagnetic and hadronic showers." Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **290**, 23 (1990) 469 – 488.

[55] Adapted from http://www-cdf.fnal.gov/physics/new/top/2004/jets/cdfpublic.html. Accessed August 2012.

[56] Eidelman, S. and others (Particle Data Group). Phys. Lett. B **592** (2004) 1.

[57] Corcella, G., et al. "HERWIG 6: An Event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)." JHEP **0101** (2001) 010.

[58] Ellis, R. K., W. J. Stirling, and B. R. Webber. "Jet properties beyond fixed order." In "QCD and Collider Physics," volume 8, pages 193–236. Cambridge University Press (1996).

[59] Abe, F. et al. "Prompt photon cross section measurement in $\overline{p}p$ collisions at $\sqrt{s}$ = 1.8 TeV." Phys. Rev. D **48** (1993) 2998–3025.

[60] Aaltonen, T. et al. "Top quark mass measurement using the template method at CDF." Phys. Rev. D **83** (2011) 111101.

[61] Freeman, J. et al. "An artificial neural network based b jet identification algorithm at the CDF experiment." Nucl. Instrum. Methods A **663**, 1 (2012) 37 – 47.

[62] Acosta, D. et al. "Measurement of the $t\bar{t}$ production cross section in $p\bar{p}$ collisions at $\sqrt{s}$ = 1.96 TeV using lepton + jets events with secondary vertex *b*-tagging." Phys. Rev. D **71**, 5 (2005) 052003.

[63] Abulencia, A. et al. "Measurement of the $t\bar{t}$ production cross section in $p\bar{p}$ collisions at $\sqrt{s}$ = 1.96 TeV using *lepton + jets* events with jet probability *b*-tagging." Phys. Rev. D **74** (2006) 072006.

[64] Acosta, D. et al. "Measurement of the $t\bar{t}$ production cross section in $p\bar{p}$ collisions at $\sqrt{s}$ = 1.96 TeV using lepton plus jets events with semileptonic *B* decays to muons." Phys. Rev. D **72** (2005) 032002.

[65] Hoecker, A., et al. "TMVA: Toolkit for Multivariate Data Analysis." PoS **ACAT** (2007) 040.

[66] Richter, S. *Search for electroweak single top-quark production with the CDF II experiment*. Ph.D. thesis, Universität Karsruhe (2007). FERMILAB-THESIS-2007-35.

[67] Ferrazza, C. *Identificazione di quark pesanti in getti adronici in interazioni $p\overline{p}$ con il rivelatore CDF al Tevatron*. Master's thesis, Università "La Sapienza" Roma (2006).

[68] Mastrandrea, P. *Study of the heavy flavour fractions in z+jets events from proton-antiproton collisions at energy = 1.96 TeV with the CDF II detector at the Tevatron collider*. Ph.D. thesis, Università degli Studi di Siena (2008). FERMILAB-THESIS-2008-63.

[69] Kim, Y. K., et al. "Comparison of quark and gluon jets produced in high-energy e^+e^- annihilations." Phys. Rev. Lett. **63** (1989) 1772–1775.

[70] Alexander, G., et al. "A comparison of b and uds quark jets to gluon jets." Zeit. Phys. C **69** (1995) 543–560.

[71] Aaltonen, T., et al. "Measurement of the top quark mass and $p\overline{p}\rightarrow t\bar{t}$ cross section in the all-hadronic mode with the CDF II detector." Phys. Rev. D **81** (2010) 052011.

[72] Aaltonen, T., et al. "Search for the Higgs boson in the all-hadronic final state using the CDF II detector." Phys. Rev. D **84** (2011) 052010.

[73] Alwall, J. et al. "MadGraph/MadEvent v4: the new web generation." J. High Energy Phys. **09** (2007) 028.

[74] Aaltonen, T. et al. "Observation of Single Top Quark Production and Measurement of $|V_{tb}|$ with CDF." Phys. Rev. D **82** (2010) 112005.

[75] Junk, T. "Building a More General χ^2." (2007). CDF/DOC/STATISTICS/PUBLIC/7904.

[76] Junk, T. "Sensitivity, Exclusion and Discovery with Small Signals, Large Backgrounds, and Large Systematic Uncertainties." (2007). CDF/DOC/STATISTICS/PUBLIC/8128.

[77] Klimenko, S., J. Konigsberg, and T. M. Liss. "Averaging of the inelastic cross sections measured by the CDF and the E811 experiments." (2003). FERMILAB-FN-0741.

[78] James, F. "MINUIT: Function minimization and error analysis." (1998). CERN Program Library D506.

[79] Feldman, G. J. and R. D. Cousins. "Unified approach to the classical statistical analysis of small signals." Phys. Rev. D **57** (1998) 3873.